

**「知の創生と情報社会」研究領域 領域活動・評価報告書**  
**－平成24年度終了研究課題－**

研究総括 中島 秀之

## 1. 研究領域の概要

本研究領域は多様もしくは大規模なデータから、有用な情報である「知識」を生産し、社会で活用するための基盤的技術となる研究を対象としています。

具体的には、大規模データを処理するための革新的な技術、統計数理科学を応用した分析・モデル化技術、あるいは実社会から得られる多様なデータを構造化・分析して知識を抽出する技術、センサによる情報取得やシミュレーション結果等の複数のリソースから新たな知識を創出する技術などの基盤技術に加えて、獲得した知識を実社会に適用するために必要とされる、シミュレーション、データの可視化、新しい情報社会の仕組みを支える応用技術などに関する研究を含んでいます。

## 2. 事後評価対象の研究課題・研究者名

件数：11件

※研究課題名、研究者名は別紙一覧表参照

## 3. 事前評価の選考方針

選考の基本的な考えは下記の通り。

1) 選考は、「知の創生と情報社会」領域のアドバイザー9名の協力を得て、研究総括が行う。

2) 選考方法は、書類選考、面接選考及び総合選考とする。

- 書類選考において1提案につき3名の選考委員が査読評価を行なう。
- 選考委員の所属機関と応募者の所属機関が異なるよう配慮し、書類選考は利害関係者を査読対象とせず、面接選考において利害関係者は席を外して実施する。
- 査読結果に基づき、3年型と5年型に分けて、事前に総括と事務局とで順位付けを施す。
- 面接選考では可能な限り多くの研究提案を直接聴取し、質疑応答する。特に5年型については、初年度の5年型に相応しい提案か否か(研究構想が本領域の趣旨に合っているだけでなく、研究期間の後半あるいは期間終了後において、実社会での応用がしっかりと考えられているかどうか)について質疑する。
- また、制度として、5年型を3年型に回すことはできなかつたため、再度挑戦の価値がある提案については、不採択理由にコメントをつけて次回以降に期待することにした。

3) 基本的に、3年型は「知の創生」の基盤技術や理論を開発するもの、5年型は実社会での適用や実運用のためのアプリケーションの開発など、「情報社会」での応用を目指すものを求めた。

## 4. 事前評価の選考の経緯

一応募課題につき領域アドバイザー3名が書類審査し、書類選考会議において面接選考の対象者を選考した。続いて、面接選考および総合選考により、採択候補課題を選定した。

選考	書類選考	面接選考	採択数			
			12件	内訳	3年型	5年型
対象数	64件	22件				

( )内は大挑戦型としての採択数。

備考:

1)平成 21 年度採択課題のうち、以下は今年度事後評価対象としない。

・伊藤孝行研究者

最先端・次世代研究開発支援プログラムに採択されたため、平成 22 年度末でさきがけ研究を中止した。

## 5. 研究実施期間

平成 21 年 10 月～平成 25 年 3 月

## 6. 領域の活動状況

### ● 領域会議

－ 7 回実施、うち 1 回は「情報環境と人」研究領域と合同。

### ● 研究成果報告会 (JST 東京本部別館) 実施 (平成 24 年 12 月 14 日)

－ 「しみわたる情報技術・にじみ出す知 ― いま、収穫の時！」をテーマに実施。

### ● 研究総括の研究実施場所訪問、他

－ 研究総括は、公立はこだて未来大学内(函館)、技術参事および事務参事は平成 20 年 6 月始～平成 23 年 11 月末まで科学技術振興機構三番町ビル内、平成 23 年 11 月末以降は科学技術振興機構東京本部別館内にて、バーチャル領域事務所態勢で業務を実施。

－ 研究場所訪問は、「サイトビジット」とも呼び、研究環境が十分であるかどうかの確認と、総括が上司の方にご挨拶することが目的。平成 21 年 10 月から 12 月にかけて、東京、藤沢、筑波、名古屋、福岡に 11 名の研究者および上司の方と面談。平成 22 年 1 月研究開始(海外から 1 月に帰国のため)の 1 名については、1 月に、研究者および上司の方と面談。研究期間中に異動となった研究者 1 名については、平成 24 年 6 月に新しい研究場所の訪問を実施。また、大学とは別に研究実施場所(スタジオ)を持つ 1 名については、平成 24 年 6 月にスタジオの訪問を実施。

－ 当領域では、研究者間のコラボレーションを重視しており、その一環として、サイトビジットの他、研究者の希望や状況把握のため、領域事務所単独での研究者訪問(レクチャービジット)を実施、平成 22 年 6 月以降、研究者が自発的に「オフ会」という交流会を開催し、積極的なコラボレーションを図っている。

### ● 他研究領域とのコラボレーション

－ さきがけ「情報環境と人」研究領域と合同で、アウトリーチ活動の一環として、情報処理学会第 72 回全国大会(創立 50 周年記念大会)にてさきがけセッションを開催(平成 22 年 3 月 11 日)、また、CREST「共生社会に向けた人間調和型情報技術の構築」領域、さきがけ「情報環境と人」研究領域と合同で、シンポジウム「情報学による未来社会のデザイン～健全でスマートな社会システムに向けて～」を企画、第一回「大量データにもとづく未来社会のデザイン」を開催(平成 24 年 11 月 8 日)、このシンポジウムは来年、再来年と 3 年連続で開催の予定。

## 7. 事後評価の手続き

研究者の作成した研究報告書および自己評価を基に、年 2 回の領域会議における経過報告および討議内容、領域アドバイザーの意見、さらに成果報告会(公開)での発表を参考にして研究総括が総合評価を行なった。

(評価の流れ)

平成 24 年 10 月	第 8 回領域会議(総括・アドバイザーによる進捗評価とアドバイス)
平成 24 年 12 月	研究報告会開催(一般参加者および総括・アドバイザーによる評価)
平成 25 年 2 月	第 9 回領域会議(総括・アドバイザーによる進捗評価とアドバイス)
平成 25 年 3 月	研究期間終了
平成 25 年 3 月	研究報告書提出
平成 25 年 3 月	研究総括による評価

備考:平成 21 年度採択課題のうち、以下は今年度事後評価を実施しない。

・伊藤孝行研究者:最先端・次世代研究開発支援プログラムに採択されたため、平成 22 年度末でさきがけ研究を中止。



## 8. 事後評価項目

- (1) 当初の研究計画から見た進捗状況や達成度
- (2) 当初計画で想定されていなかった新たな展開が生じているか否か
- (3) 成果の科学的・技術的インパクト、国内外からの類似研究と比較したレベルや重要度

## 9. 事後評価

第二期生は当初 12 名で研究開始したが、途中、1 名が最先端・次世代研究開発支援プログラムに採択され、修了時は 11 名であった。

北本研究者の「ベイジアンテレビ:取材・配信・編集を自動化した緊急情報メディア」は、東日本大震災が発生し、実際の緊急情報を扱うことになり、急遽、研究計画を変更して震災関連情報を取り扱うことで、文化庁のメディア芸術祭で表彰された。また、高田研究者「金融市場における相転移の時空間構造の自動抽出と予測」は、本領域の情報処理専門家とのコラボにより、当初予定の金融情報の分析から社会科学的研究に踏み込んだ成果をえることができた。その他の研究者も、ほぼ当初研究計画を達成している。今期は成功であったと考える。

### 1. 赤石美奈研究者「物語構造に基づく情報編纂基盤技術」

文書中の単語の共起依存関係だけから文書に書かれた文脈(物語)を抽出し、DB 化や新規文書構成、文脈の検索などを可能にすることを旨とした研究課題であった。

連想情報アクセス手法の情報探索手法としての研究について、いくつかの応用分野において成果をあげている。「物語」に関する編纂基盤技術として、さらに開発された基盤技術を統合環境として提供できるようにしていくのがよいであろう。

今後、文学作品等に適用し、物語生成支援へと研究を展開させていってほしい。

### 2. 上野玄太研究者「次世代データ同化:自動モデル化と情報フロー抽出技術開発」

データとモデルを組み合わせることにより、モデルの高性能化(予測・推定)技術、モデル作成の省コスト化技術、モデルが捉えた情報の流れの抽出技術を系統的に開発するという研究課題であった。「省コストのデータ同化手法の開発」、「モデルが捉えた情報の流れの抽出技術」、「モデルの高性能化技術」の3テーマに分けて取り組んでいる。

「省コストのデータ同化手法の開発」と「モデルが捉えた情報の流れの抽出技術」については、シミュレーションモデルをもとにしたデータ同化システムやアンサンブル微分法等の開発により解決している。「モデルの高性能化技術」については課題達成に至らなかったが、その過程で得られた、モデルの高性能化のために、説明すべき変数と入力変数を明確に抽出できたこと、それらのモデル化が必要であるという認識が得られており、有意義な成果と考えられる。今後の継続研究に期待したい。

### 3. 宇野毅明研究者「大規模データに対する高速類似性解析手法の構築」

大規模データ内の汎用な事例間類似性高速計算アルゴリズムの開発が研究課題である。

今回の研究で、文字列を中心として、非常に短い時間で、今までは考えられなかったような長いパターンをたくさん見つけるアルゴリズムを開発することに成功し、長い頻出文字列パターンのマイニングが初めて可能となっている。また、これを文字列の圧縮アルゴリズムにも適用し、最適な選択を行うアルゴリズムの開発に成功している。

また、研究交流会(オフ会)を主催し、研究者間の新たな共同研究を生み出す等、領域内のコラボレーションに大きく寄与しており、高く評価している。

今後、他の類似性尺度などの関係性についても手法の研究の範囲を拡大していってほしい。また、実際のデータ解析についても研究を展開していってほしい。

### 4. 岡部誠研究者「映像分析による知識の抽出と、その利用による新たな映像合成」

動画 DB を参照しながら静止画を動画として合成する研究である。

大量の映像データを統計的に分析して知識を抽出し、ユーザが直感的な映像製作を可能にする環境を作ることがテーマである。このテーマに沿って、「ビデオデータベースを用いて流体画像をアニメーションさせる手法」、「動画鑑賞のためのスケッチインタフェース」、「入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術」、「アマチュア演技者の動作を格好良く見せるように動画を編集する手法」、「勾配画像処理に基づく動画からの流体抽出」等の基本的な技術の研究開発を行った。それぞれ、「目で見て分かる成果

をあげている。

今後、このさきがけ研究を通じて得られた、2次元動画の解析・学習の限界とコンピュータに人間と同様に3次元環境を学習させるべきとの知見に基づき、2次元流体動画の応用研究を3次元流体動画に拡張していくと同時に、これらの成果を統合して、どのような方向に向かうのかを明確にしながら研究を進めてほしい。

#### 5. 北本朝展研究者「ベイジアンテレビ：取材・配信・編集を自動化した緊急情報メディア」

利用者が情報を引き出す行為がなくても重要な情報が利用者に届くようなプッシュ型メディアの構築を目標とし、取材・配信・編集を自動化した「ベイジアンテレビ」を開発した。

当初は台風関連データを対象としていたが、途中で東日本大震災が発生し、大震災関連データを対象に加えて事態の推移にリアルタイムで対応する取り組みに方針変更し、データを活用したサービスも開発している。想定外の多種で大量なデータであったために、一時、研究が遅延したが、データの処理方式や表示方式の見直しにより、それを挽回している。入力と出力をつなぐ部分、具体的には情報源と利用者をつなぐ「情報の交換機」となる部分は、課題未達成であるが、メディアアートとして文化庁メディア芸術祭で受賞等、各種の賞を獲得しており、高く評価している。

「テレビ」は最大にして最強のメディアとのことで、テレビを最終目標とした計画を維持しているが、今後はテレビというメタファーに囚われず、自由な発想で研究を進めてもらいたい。

#### 6. 坂本比呂志研究者「圧縮データ索引に基づく巨大文書集合からの関連性マイニング」

文字列共起を利用する超大規模なテキストなどの生文書の類似性計算、文書分類、検索に関する新しい技術の開発がテーマである。

さきがけ研究を通じ、文法圧縮の理論をほぼ完成させ、プログラムを公開して他の研究者に提供し、応用例を通して今後の可能性を示しており、評価できる。文書群からのマイニングと並列化による大規模化は果たしていないが、データを圧縮したままでの情報処理の道を切り開いており、評価できる。

「圧縮情報学」を提唱し、この分野を発展させたいとしており、今後に期待している。

#### 7. 佐久間淳研究者「実社会情報ネットワークからのプライバシー保護データマイニング」

プライベートなデータを扱うデータマイニングにおいて、プライバシーを確保するプライベート・グラフ・モデルの提案、およびその手法についての研究である。

ネットワーク環境でプライバシーを安全に利用するという点について、プライベートグラフという基礎概念を元に、主にネットワークマイニングや多人数意思決定のための多くのアルゴリズムをマルチパーティー秘密計算として実行するためのプロトコル群を提供したことは評価できる。

また、現実世界でプロトコルを動作させ、またプロトコル開発を容易にするためのフレームワークである Fairy Ringを開発し、これを利用して様々な応用を見いだす体制を開発しており、今後に期待したい。

#### 8. 杉山将研究者「密度比推定による大規模・高次元データの知的処理技術の創生」

データ解析の分野で、密度比推定という枠組みをさまざまな応用分野に適用可能にするという研究である。

この研究では、密度比推定という新技術を軸に、密度比を用いることによって難しかった問題を簡単な形に変換して解くというアプローチで、その技術が応用できる実問題を探するというスタイルで研究を進め、多くの企業と共同で多数の応用事例を解決している。

この技術を普及させるための教科書も出版している。着実に成果をあげていることを評価したい。また、理論と応用の橋渡しを進めていくため、基礎から応用までをカバーする研究チームを作っていくとのことであり、期待している。

#### 9. 鈴木秀幸研究者「インフルエンザ感染伝播のデータ同化モデルによる解析・予測技術」

データ同化技術の導入でインフルエンザ感染伝播のモデルを構築し、解析・予測を行うための数理的基盤技術を開発するという研究であった。

結果的にはメタポピュレーションモデルなどによる既存の感染症モデルに基づく研究となった。また、感染症の多様性を考慮した数理モデル研究と、データ同化研究とをまとめきれていないようであるが、当初の予定であった感染症伝播のデータ同化シミュレーションに関する成果を得ている。

今後、感染症の多様性を考慮した数理モデル研究と、データ同化研究の融合をさらに進めてほしい。

#### 10. 高田輝子研究者「金融市場における相転移の時空間構造の自動抽出と予測」



金融・経済データ情報から金融バブルなどの相転移現象の時空間構造を可視化し、相転移メカニズムの探索・モデル化と予測を可能にする知識生産方法を開発し、金融・経済のシミュレーションモデルの新しいデータ同化技術構築の基盤技術を目指すという研究課題である。

当初目標は、金融市場における相分類による相転移予測を大規模金融データとロバスト・効率的手法の開発により、実用レベルの精度に高めることであったが、相転移研究のための技術的困難を、統計手法とデータの大規模化の両面から、ほぼ解消することができている。また、さきがけの特徴である研究者間のコラボレーションをうまく活用し、当初の研究範囲を超え、日米の売手/買手別指値表による板値の最長時系列データやウェブ上の大規模投資家掲示板最長時系列データ、格付けが利用できる米国上場のほぼ全ての社債データ等々、他種の大規模データの収集・解析を行い、相転移予測の精度を高めることができたことを評価する。今後、他の複雑系における相転移現象で得られている知見の活用や、投資家心理のダイナミクスの分析等を進め、金融バブルの真実に迫り、社会に役立つ研究成果に結実させて行ってほしい。

#### 11. 中西泰人研究者「空間的な情報システムの設計開発支援システム」

デザインプロセスをモデル化し、それに基づいた設計支援システムの構築が課題である。

研究期間中、空間的な情報システムを設計開発する方法を複数提案し、適用・評価し、設計開発支援システムに機能を追加してバージョンアップを繰り返している。広汎な都市空間を対象とした空間的な情報システムの運用には至らなかったとはいえ、ユーザを一般公募してワークショップを開催し、ドキュメント整備も行き、こんごの普及体制を整えたことを評価したい。

今後、映像装置だけでなく、他のコミュニケーションモードを設計開発の対象とできるようにシステムを拡張し、広汎な都市空間を対象とした空間的な情報システムに取り組んでいってほしい。

#### 10. 評価者

研究総括 中島 秀之 公立はこだて未来大学・学長

領域アドバイザー(五十音順。所属、役職は平成25年3月末現在)

麻生 英樹 産業技術総合研究所知能システム研究部門・主任研究員  
 有村 博紀 北海道大学大学院情報科学研究科・教授  
 高野 明彦 国立情報学研究所連想情報学研究開発センター・センター長／教授  
 林 晋 京都大学大学院文学研究科・教授  
 林 幸雄 北陸先端科学技術大学院大学知識科学研究科・准教授  
 樋口 知之 統計数理研究所・所長  
 堀 浩一 東京大学大学院工学系研究科・教授  
 安田 雪 関西大学社会学部・教授  
 鷲尾 隆 大阪大学産業科学研究所・教授

(参考)

件数はいずれも、平成25年3月末現在。

##### (1) 外部発表件数

	国内	国際	計
論文	25	67	92
口頭	108	78	186
その他	12	11	23
合計	145	156	301

##### (2) 特許出願件数

国内	国際	計
6	6	12

### (3)受賞等

- 宇野毅明研究者
  - ・ 文部科学大臣表彰 科学技術部門 若手科学者賞 (H22.4)
- 岡部誠研究者
  - ・ 優秀ポスター賞 (H22.12)
  - ・ DEIM 2012 学生奨励賞 (H24.3)
  - ・ VC 賞 (H24.6)
  - ・ グラフィクスと CAD 研究会優秀研究発表賞 (H24.6)
- 北本朝展研究者
  - ・ CSIS Days 2011 優秀研究発表賞 (H23.11)
  - ・ 第 16 回文化庁メディア芸術祭アート部門審査委員会推薦作品(H24.12)
  - ・ Linked Open Data Challenge 2012 ビジューライゼーション部門優秀賞(H25.3)
  - ・ オープンデータ流通推進コンソーシアム勝手表彰 ソフトバンクテレコム賞(H25.3)
- 佐久間淳研究者
  - ・ 第 23 回人工知能学会全国大会優秀賞 (H21.11)
  - ・ 情報論的学習理論と機械学習(IBISML2010)研究会 Honorable Mention (H22.11)
  - ・ 情報論的学習理論と機械学習(IBISML) 2010 年度研究会賞 (H23.6)
  - ・ 平成 24 年度 日本データベース学会上林奨励賞 (H25.3)
- 杉山将研究者
  - ・ 電子情報通信学会 PRMU 研究会 研究会奨励賞 (H22.5)
  - ・ 人工知能学会 DMSM 研究会 研究会優秀賞 (H22.5)
  - ・ 情報処理学会 平成 22 年度長尾真記念特別賞 (H23.6)
  - ・ 2011 年度日本神経回路学会論文賞 (H23.12)
  - ・ 船井情報科学振興財団 船井学術賞 (H24.4)

### (4)招待講演

国際 11 件

国内 6 件

研究者氏名 (参加形態)	研究課題名 (研究実施場所)	現職(平成25年3月末現在) (応募時所属)	研究費 (百万円)
赤石 美奈 (兼任)	物語構造に基づく情報編纂基盤技術 (法政大学情報科学部)	法政大学情報科学部コンピュータ 科学科・教授 (東京大学大学院工学系研究科・ 准教授)	38
上野 玄太 (兼任)	次世代データ同化:自動モデル化と情 報フロー抽出技術開発 (統計数理研究所)	統計数理研究所モデリング研究系・ 准教授 (同上・助教)	37
宇野 毅明 (兼任)	大規模データに対する高速類似性解 析手法の構築 (国立情報学研究所)	国立情報学研究所プリンシプル研 究系・准教授 (同上)	39
岡部 誠 (兼任)	映像分析による知識の抽出と、その利 用による新たな映像合成 (電気通信大学 情報理工学部)	電気通信大学 情報理工学部 総 合情報学科・助教 (マックスプランク情報科学研究所・ 博士研究員)	31
北本 朝展 (兼任)	ベイジアンテレビ:取材・配信・編集を 自動化した緊急情報メディア (国立情報学研究所)	国立情報学研究所コンテンツ科学 研究系・准教授 (同上)	39
坂本 比呂志 (兼任)	圧縮データ索引に基づく巨大文書集 合からの関連性マイニング (九州工業大学大学院情報工学研究 院)	九州工業大学大学院情報工学研究 院・准教授 (同上)	36
佐久間 淳 (兼任)	実社会情報ネットワークからのプライ バシ保護データマイニング (筑波大学大学院システム情報工学研 究科)	筑波大学大学院システム情報工学 研究科・准教授 (同上)	37
杉山 将 (兼任)	密度比推定による大規模・高次元デー タの知的処理技術の創生 (東京工業大学大学院情報理工学研 究科)	東京工業大学大学院情報理工学研 究科・准教授 (同上)	39
鈴木 秀幸 (兼任)	インフルエンザ感染伝播のデータ同化 モデルによる解析・予測技術 (東京大学生産技術研究所)	東京大学生産技術研究所・准教授 (同上)	21
高田 輝子 (兼任)	金融市場における相転移の時空間構 造の自動抽出と予測 (大阪市立大学大学院経営学研究科)	大阪市立大学大学院経営学研究 科・准教授 (同上)	39
中西 泰人 (兼任)	空間的な情報システムの設計開発支 援システム (慶應義塾大学環境情報学部)	慶應義塾大学環境情報学部・准教 授 (同上)	34

# 研究報告書

## 「物語構造に基づく情報編纂基盤技術」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 赤石 美奈

### 1. 研究のねらい

現代の知識化社会において、人々は膨大な情報の中から、それぞれが抱えている問題に関連する情報を収集・分析し、迅速かつ的確にまとめて、己自身や所属する組織等の必要とする知識を獲得して意志決定に供することが求められている。

検索エンジン等で代表される情報検索技術は、利用者が必要とする情報が、(文書のような)検索単位として、予め存在しているような場合には、有効な検索手段である。しかしながら、利用者の要求を満足する情報が存在するかどうか分からない場合、単一の文書では情報要求に応えられず複数の文書の組み合わせが必要な場合、あるいはそもそも利用者自身が情報要求を分節できていないような場合には必ずしも有効であるとはいえない。

このような情報要求を満たすには、情報を検索しては解釈し、また検索するといった試行錯誤が必要となる。このような探索的情報アクセスを支援するためには、試行錯誤の際に、より有用な方向へ進めるように、情報洪水の中に道標を示すことが必要であろう。

本研究では、「物語」に着目し、大量な情報の中に潜在する多重・多層文脈を顕在化し、多様な視点や文脈に応じて有益な情報を編纂することを支援するシステムの構築を目指す。これにより、知識創造につながる探索的情報アクセスを支援する情報アクセス環境の実現を図る。本システムは、単にユーザが直面している問題の解答を探索するのみではなく、場合によっては、問題や要求そのものが変化し、その解決法を探索するプロセスにより、新しい概念世界を創造することを目指している。このために、既存の情報を発掘するだけではなく、新しい物語を生成するための機構を実現しなければならないと考えている。つまり、検索技術が(過去に生成された)蓄積物を対象にしているのに対して、ナラティブ連想情報アクセスでは、これから生成される新しい物語を産出す技術の研究開発を目指している。

### 2. 研究成果

#### (1) 概要

本研究では、文書の持つ『物語構造』に着目し、大量な情報の中に潜在する多重・多層文脈を顕在化し、多様な視点や文脈に応じて有益な情報を編纂することを支援するシステムの構築を目指した。これにより、知識創造につながる探索的情報アクセスを支援する情報アクセス環境の構築を進めた。

基盤技術としては、物語構造に基づき、新しい文脈を生成しながら、物語を生成するナラティブ連想情報アクセスのフレームワークについて研究・開発を行った。まず、文書における物語構造モデルを定義し、ナラティブ連想情報アクセス・フレームワークの基本を提案し、「語の共起依存性に基づく語彙連鎖構造」を基に、基盤技術となる文書分割方法、断編連結方法の

研究を進め、応用システムであるナラティブ連想情報アクセス・システムの開発に関する研究に取り組んだ。

ナラティブ連想情報アクセスのフレームワークは、新しい文脈を生成しながら、物語を生成するための枠組みである。基本的な構造は、文脈に応じて、既存の知識を再構築するために、既存の文書を分割し、再構成する仕組みである。

さらに、このフレームワークに、概念空間の概念を導入し、物語の表面には明示されていない関係を見つけるための仕組みについて研究し、潜在文脈の顕在化に焦点をあてて、研究を進めた。一つの文脈に沿って表現されている文書群を別の観点から捉えなおすことで、新しい文脈に基づく物語が顕在化される。

また、概念空間に、物語の構成場面を連結するための場のポテンシャルと連結力に加わる外力の導入を検討し、概念空間の変化により物語構造を変形される仕組みについても研究を行った。

## (2) 詳細

### 研究テーマ「ナラティブ連想情報アクセスのフレームワークの構築」

本研究では、物語構造に基づき、新しい文脈を生成しながら、物語を生成するナラティブ連想情報アクセスのフレームワークについて研究を行った。本研究では、「語の依存性に基づく語彙連鎖構造」を基に、基盤技術となる(i)文書分割方法、(ii)断編連結方法の研究を進め、応用システムであるナラティブ連想情報アクセス・システムの開発に関する研究に取り組んだ。

#### (i) 文書分割方法: 語彙連鎖に基づく物語分割

ひとつの文書には、複数の主題が含まれている。文脈に応じた再構成を考えた場合、既存の文書を主題毎の断編に分割する必要がある。本研究においては、「語の共起依存性に基づく語彙連鎖構造」に基づく系列的文書分割法と場面的文書分割法などの文書分割手法を開発した。

#### (ii) 断編連結法: 語彙連鎖に基づく文脈生成

本研究で提案するナラティブ連想情報アクセスは、対象文書を、物語構造に基づき分節し、ユーザが選択した文脈に沿って生成される新しい物語の候補を結果として出力し、既存文書集合を横断的に再構成して得られる新しい知識獲得を支援するものである。本研究においては、各種の文書集合に対して、トピック遷移パターンの解析を行い、応用分野において適切な物語生成を可能とする遷移ルールの確立と、支援システムの構築を目指した。

### 研究テーマ「動的に変化する概念空間の構築」

多様な物語を生成するための母体となる概念空間は、固定化されるものではなく、状況に応じて変化可能な仕組みを持たせる必要がある。状況に応じて、拡張・収縮可能な概念空間の設計を行った(論文 4 参照)。

### (i) 概念語ベクトルに基づく概念空間

概念空間の基本軸となる概念語ベクトルを動的に生成することにより、物語構造が動的に変化できる概念空間を実現する。図 1 は、同一の歴史史料を異なる概念語の配置された概念空間にて配置した例である。

### (ii) 時間軸と概念語ベクトルを軸とする概念空間

時間変化が重要である編年型データにおいては、時間軸を基にした概念空間が必要である。これに特化した概念空間の特性に関して研究を進め、多様な物語を生み出す仕組みの基盤を形成した。図 2 は、時間軸と概念語(人名)に基づくデータ分布を示す。このような概念空間から、一連の関連情報を抽出し、歴史解釈を展開させる基本プロットの構築を図った(論文 2, 3 参照)。

## 研究テーマ「潜在文脈の顕在化」

### (i) 潜在的な物語を抽出するための潜在文脈の顕在化

概念空間に配置された物語構成場面間の関連を基に、潜在する文脈を顕在化する手法について研究した。各構成要素場面同士の連結は、語の吸引力遷移パターンとして定義される。図 3 は、人工衛星の設計議事録を用い、衛星の運用上の問題解決のための情報へアクセス可能なことを示した Topic Tracer システムである。図 4 は、様々な視点に基づき形成される概念空間から抽出される一連の関連情報の例である。ここでは、人間関係の変遷が、視点となる概念や人物によって異なることを示している。図 5 は、プレゼンテーションに利用されたスライドを語の連想パターンに基づき、再構成し、新たな文脈におけるプレゼンテーション資料作成のための支援ツールを示す。

(ii) 物語生成のためのミッシングリンクの補間  
概念空間内の物語構成場面のみでは、有意な文脈を形成できない場合がある。その際に、概念空

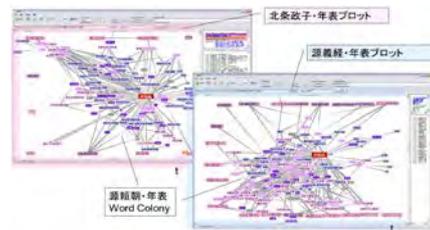


図 1. 多視点に基づく歴史史料の捉え方

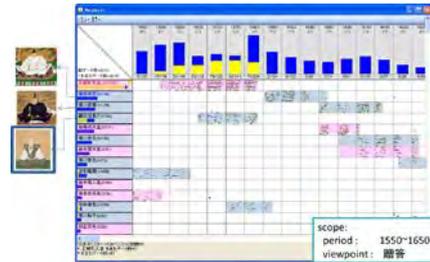


図 2. 時間軸と概念語の出現例

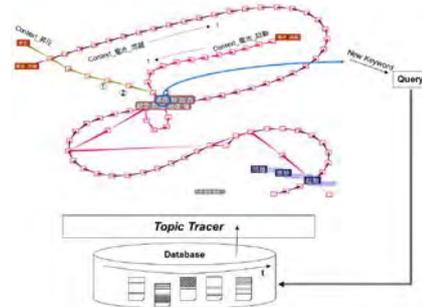


図 3. 設計議事録からの知識獲得



図 4. 多視点からの人間関係の変遷

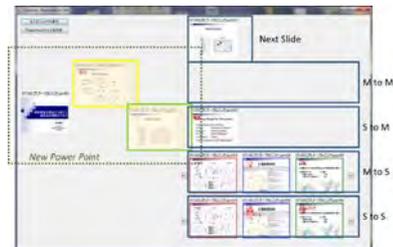


図 5. プレゼンテーションの動的再構成

間の構造自体を変化させ、物語構造を形成する方法と、不足している場面情報を概念空間内に取り込み、物語構造を形成する方法が考えられる。これらの場面連結の欠損(ミッシングリンク)を補間する技術について、トピックブリッジング手法として研究を進めた(論文 1, 5 参照)。

以上、様々な対象に対して、ナラティブ連想情報アクセスのフレームワークを適用し、応用可能なことを示した。

### 3. 今後の展開

本研究においては、「物語」を意図的に関連付けられた一連の出来事としてとらえ、「編集」を出来事間の関連付けることとしてきた。動的な物語編集技術とは、状況に応じた物語(知識)の生成技術を意味し。これにより、視点を少しずつ変化させながら関連情報を辿る技術の研究を進めてきた。

今後は、実際の文学作品としての物語生成支援へと研究を展開させていく予定である。このためには、一見、関連のなさそうな概念を結び付けたり、想像していなかった結果へと物語を展開させていく編纂技術が必要となる。また、さらに日常生活で得られる情報(ライフログ)から生成される物語を通じ、相互に影響しあう、現実世界と情報空間から物語を生み出す仕組みを考案していきたいと考えている。

### 4. 自己評価

情報探索手法としてのナラティブ連想情報アクセス手法の研究に関しては、いくつかの応用分野において研究を進めることができた。しかしながら、未だ、「物語」に関する編纂基盤技術としては、まとめきれておらず、開発された基盤技術を統合環境として提供するには至らなかった。また研究成果の発表が滞っており、今後、整理して、まとめあげることが必要である。

### 5. 研究総括の見解

文書中の単語の共起依存関係だけから文書に書かれた文脈(物語)を抽出し、DB 化や新規文書構成、文脈の検索などを可能にすることを目指した研究課題であった。

連想情報アクセス手法の情報探索手法としての研究について、いくつかの応用分野において成果をあげている。「物語」に関する編纂基盤技術として、さらに開発された基盤技術を統合環境として提供できるようにしていくのがよいであろう。今後、文学作品等に適用し、物語生成支援へと研究を展開させていってほしい。

### 6. 主な研究成果リスト

#### (1) 論文(原著論文)発表

1. Makoto Sato, Mina Akaishi, Koichi Hori: Topic Bridging by Identifying the Dynamics of the Spreading Topic Model, Intelligent Interactive Multimedia: Systems and Services, Proceedings of the 5th International Conference on Intelligent Interactive Multimedia Systems and Services (IIMSS 2012), Smart Innovation, Systems and Technologies, Vol. 16, pp. 619-628, 2012.

2.	<u>Masahiko ITOH</u> , <u>Mina AKAISHI</u> , Visualization for Changes in Relationships between Historical Figures in Chronicles, Proc. of Int. Con. on Information Visualisation 2012, pp.283-290, 2012
3.	<u>Mina Akaishi</u> , Makoto Sato, <u>Koichi Hori</u> , Tetsuya Ishikawa: A Visual Analysis Tool for Finding Embedded Relations in Chronicles, In Proceeding of 16th International Conference on Information Visualisation (IV), pp.302-307, 2012.
4.	<u>Mina Akaishi</u> , Makoto Sato and <u>Koichi Hori</u> : A Visual Analysis Tool for Amplifying Story Generation Cycle, Proc. Of IADIS International Conference, Computer Graphics, Visualization, Computer Vision and Image Processing 2011, pp.349-352, 2011
5.	Makoto Sato, <u>Mina Akaishi</u> , <u>Koichi Hori</u> : Topic Bridging for Story Generation, Information Modelling and knowledge Bases XXII, pp.247-257, 2011

(2)特許出願

研究期間累積件数:0件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

# 研究報告書

## 「次世代データ同化:自動モデル化と情報フロー抽出技術開発」

研究タイプ:通常型

研究期間:平成21年10月～平成25年3月

研究者:上野 玄太

### 1. 研究のねらい

大規模データから有用な知識を得るためには、よいモデルを立てること、さらにモデルを通して得た情報の中から必要なものを抽出する必要がある。本提案は、データとモデルを組み合わせるデータ同化手法を基盤とし、

研究テーマ A「省コストのデータ同化手法の開発」

研究テーマ B「モデルが捉えた情報の流れの抽出技術」

研究テーマ C「モデルの高性能化(予測・推定)技術」

を系統的に開発するものである。これらにより、従来は専門家の経験が必要とされていたモデルのチューニングや実行環境に応じた再モデル化、入力データ量を凌ぐ大量のモデル出力の動的な解析が可能になる。

### 2. 研究成果

#### (1)概要

研究テーマ A「省コストのデータ同化手法の開発」

データ同化の中心部であるシミュレーションモデルは、微視的な物理・化学法則に基づいて構成されていることが多く、データ同化のために多数のメンバーからなるアンサンブルを構成することは、計算機能力の観点から難しい。アンサンブルメンバーが少ないことによる難点は、メンバー間の相関を求めた場合に、見かけ上の偽の相関値が得られるという点である。そこで、そういった偽相関を抑える正則条件を組み込んだデータ同化手法を開発することとした。ここでの正則条件として、シミュレーションモデルの変数間の依存関係に着目し、アンサンブルメンバー間の相関行列の逆行列に明示的に構造を入れることとした。この構造をもとにして、逆行列を基準としたデータ同化手法(グラフベース・アンサンブルカルマンフィルタ)を開発した。

研究テーマ B「モデル変数の感度計算法の開発」

データ同化が完了しても、具体的にどの変数がどのように機能しているかを見出す作業は依然として容易ではなく、専門家の知識が必要となるのが現実である。しかし、変数の関係はすべて既知であるのだから、原理的にはモデル内に解答があるはずである。そこで、データ同化による推定値に対するモデル変数の感度を得る方法を開発することとした。微分係数のアンサンブル近似の定式化を行い、各時間ステップでの微分係数値を求める方法を考案した。巨大次元の変数への対処が課題であったが、特異値分解等の線形計算を組み合わせることで、アンサンブルメンバー数だけの手間で済むことを示した。

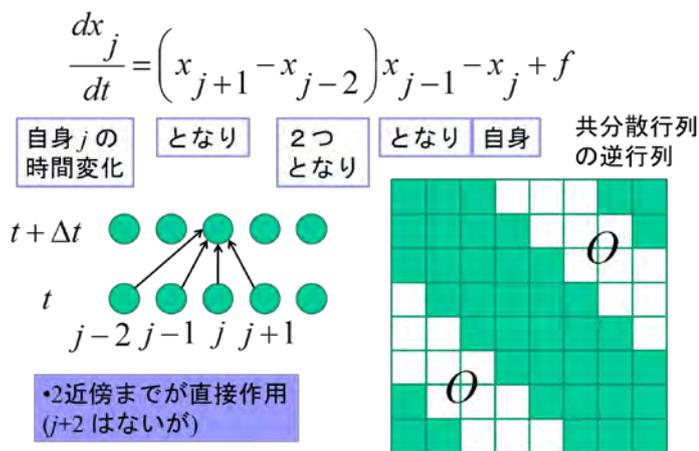
#### (2)詳細

研究テーマ A「省コストのデータ同化手法の開発」

データ同化におけるシステムモデルの核となるのは、モデル変数の時間更新を表現する部分である。すなわち、あるタイムステップにおける変数を与えたもとの、次ステップでの変数値を推定する部分である。この時間更新モデルには、自己回帰モデルなどの応用分野をそれほど意識しないものも使われるが、応用分野での知見にもとづき開発が進められてきたシミュレーションモデルがあるならば、その分野における有用性や予測能力という見地から、そういったシミュレーションモデルを時間更新部分としてそのまま用いることが望ましい。しかし、それらのモデルは微視的な物理・化学法則に基づいて構成されていることが多く、データ同化のために多数のメンバーからなるアンサンブルを構成することは、計算機能力の観点から難しくなる。

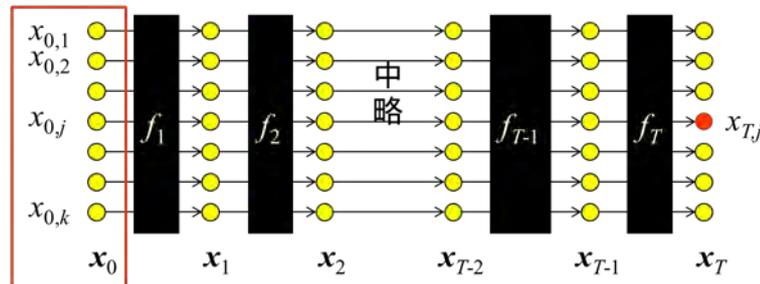
アンサンブルメンバーが少ないことによる難点は、メンバー間の相関を求めた場合に、見かけ上の偽の相関値が得られるという点である。そこで、そういった偽相関を抑える正則条件を組み込んだデータ同化手法を開発することとした。従来は、モデルの構造には着目せず、単に偽相関をテーパリングすることで抑える方法がとられているが、テーパリングの強度を調整するパラメータの選択に難点があることが知られていた。

そこで、正則条件として、シミュレーションモデルの変数間の依存関係に着目し、アンサンブルメンバー間の相関行列の逆行列に明示的に構造を入れる、新しいデータ同化手法を開発した。この方法により、モデル構造を直接反映した正則化が可能になり、従来のアドホックなテーパリングが不要となった。グラフ構造を仮定したアンサンブルカルマンフィルタととらえることができるので、グラフベース・アンサンブルカルマンフィルタ (Graph-based ensemble Kalman filter, GEnKF) と呼ぶこととする。小規模なシミュレーションモデル(40 変数)をもとに、GEnKFの実装を行い、従来の同化手法(アンサンブルカルマンフィルタ、粒子フィルタ)よりも少ないメンバー数でも真値に近い推定値が得られることを確認した。



#### 研究テーマ B「モデル変数の感度計算法の開発」

データ同化の実施により、モデルの枠組みで現象を理解することになる。これは、モデルに含まれる変数の関係が既知であるという点でデータのみの状況よりも大きく前進しているが、具体的にどの変数がどのように機能しているかを見出す作業は依然として容易ではなく、専門家の知識が必要となるのが現実である。しかし、変数の関係はすべて既知であるのだから、原理的にはモデル内に解答があるはずである。



**問題** 最終状態  $x_T$  の  $j$  成分 ( $x_{T,j}$  と書く) に多く寄与するのは、初期状態  $x_0$  のどの成分か (もしくは成分群か) ?

そこで、データ同化による推定値に対するモデル変数の感度を得る方法を開発し、モデル内の情報の流れを可視化することを考える。モデルは通常、巨大次元である上に数多くの演算をし、非線型のプロセスを含むため、単純な数値微分では有効な感度を得ることができない。アジョイント法もしくは自動微分と呼ばれる計算技術が一部で使われているが、そのコーディングと保守の困難さが問題である。

そこで、複数の異なる入力変数を与えたモデルからなるアンサンブルを用いることで、感度の近似計算法を開発した。基本的なアイデアは、状態のアンサンブルを用いて数値微分を各時間ステップで計算し、それを微分のチェーンルールに照らして掛け合わせていくというものである。具体的には、微分係数のアンサンブル近似の定式化を行い、各時間ステップでの微分係数値を求める方法を考案した。巨大次元の変数への対処が課題であったが、特異値分解等の線形計算を組み合わせることで、アンサンブルメンバー数だけの手間で済むことを示した。目標となる最終時点  $T$  の特徴的な変数 (図で示す  $x_{T,j}$ ) から、時間をさかのぼる形式 ( $T-1, T-2, \dots, 2, 1, 0$ ) で感度が逐次的に求められる。課題Aと同じ小規模な 40 変数のシミュレーションモデルに対して検証を行い、8 メンバー程度のアンサンブルでも妥当な感度が得られることを示した。

### 3. 今後の展開

研究テーマ A, B で開発したグラフベースアンサンブルカルマンフィルタ (GEnKF)、アンサンブル感度計算法を組み込んだ、大規模なシミュレーションモデルをもとにしたデータ同化システムを構築する。一つは、気象庁による数値予報モデル (非静力学モデル) を用いたシステムを考えている。計算負荷が比較的小さい (50 万変数) 設定での、温帯性低気圧の予報から始めたい。もう一つは、気候変動のアンサンブル予測である。GEnKF により、少ないメンバーでの分布表現が可能になるため、頻度は小さいが被害が甚大な事象の予測を効率的に行うことができるとともに、アンサンブル感度計算法により、温暖化の原因変数の推定を少ない手間で行うことができる。

研究途中段階であるテーマ C については、モデルで説明できない部分と入力変数のモデル化について、引き続き考察と実験を進めていきたい。

### 4. 自己評価

研究テーマ A、B に関しては、GENKF およびアンサンブル微分法の開発により、解決されたものと考えている。

研究テーマ C については、もとより容易ではない課題であると認識していたが、研究途上段階で期間の終了を迎えてしまった。ただし、その過程で得られた、モデルの高性能化のために、説明すべき変数と入力変数を明確に抽出できたこと、それらのモデル化が必要であるという認識が得られたことは有意義であると考えている。データそのもの、モデル出力そのものを考えるのではなく、誤差の影響も考慮したそれらの組み合わせから得られる変数を出発点とした、第 2 ステージのモデル化対象を明確にできたものと考えている。

## 5. 研究総括の見解

データとモデルを組み合わせることにより、モデルの高性能化(予測・推定)技術、モデル作成の省コスト化技術、モデルが捉えた情報の流れの抽出技術を系統的に開発するという研究課題であった。「省コストのデータ同化手法の開発」、「モデルが捉えた情報の流れの抽出技術」、「モデルの高性能化技術」の3テーマに分けて取り組んでいる。

「省コストのデータ同化手法の開発」と「モデルが捉えた情報の流れの抽出技術」については、シミュレーションモデルをもとにしたデータ同化システムやアンサンブル微分法等の開発により解決している。「モデルの高性能化技術」については課題達成に至らなかったが、その過程で得られた、モデルの高性能化のために、説明すべき変数と入力変数を明確に抽出できたこと、それらのモデル化が必要であるという認識が得られており、有意義な成果と考えられる。今後の継続研究に期待したい。

## 6. 主な研究成果リスト

### (1) 論文(原著論文)発表

- |  |
|--|
| 1. Ueno, G. and N. Nakamura, Iterative algorithm for maximum likelihood estimation of observation error covariance matrix for ensemble-based filters, Quarterly Journal of the Royal Meteorological Society, in press. |
| 2. 上野玄太、データ同化、システム制御情報学会誌、印刷中。   |
| 3. 上野玄太、データ同化における大規模グラフィカルモデルの推定について、統計数理、印刷中。   |
|  |
|  |

### (2) 特許出願

研究期間累積件数: 0 件

### (3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. Ueno, G., and T. Tsuchiya, Regularization of error covariance with the Gaussian graphical model, 14th Symposium on Integrated Observing and Assimilation Systems



- for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), The 90th American Meteorological Society Annual Meeting (Atlanta), January 20, 2010.
2. Ueno, G., An iterative algorithm for estimating the observation error covariance matrix for ensemble-based filters, EGU General Assembly 2011 (Vienna), April 8, 2011.
  3. Ueno, G., Derivative-Free Estimation of the Observation Error Covariance Matrix for Ensemble-Based Filters, 16th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), The 92nd American Meteorological Society Annual Meeting (New Orleans), January 25, 2012.
  4. Ueno, G., A Bayesian approach for estimating stable observation noise covariance, EGU General Assembly 2012 (Vienna), April 26, 2012.
  5. Ueno, G., Stable Estimation of Observation Error Covariance Matrix by Bayesian Method, 17th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), The 93rd American Meteorological Society Annual Meeting (Austin), January 8, 2013.

# 研究報告書

## 「大規模データに対する高速類似性解析手法の構築」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 宇野 毅明

### 1. 研究のねらい

巨大データを解析するには様々な問題がある。データが初等的でモデル化が難しく、データを解析しようと思っても、何ができるか、何をすればよいのか分からないことが多い。そのまま可視化をする、そのまま知識発見を行うのは一般に難しい。一方で、近年盛んに解析されているネットワークなどのデータでは、物と物との関係性に着目した解析が行われている。このような解析では、データのサンプリングを行ってしまうと、関係が失われてしまうことが多く、精度の高い解析が不可能となる。また、サンプリングをすると、比較的小さなまとまりはノイズと区別できなくなってしまう、ということもある。このような点から、巨大データをサンプリングして小さくしてから解析するという方法は有効でない。巨大データの関係性を直接調べ、それを元にした解析手法を実現することが肝要である。

本研究では、関係性の中でも比較的取り扱いが容易な「類似性」に着目する。一つ目の研究課題は、大規模データの類似性を網羅的に調べる効率良いアルゴリズムを構築することである。データが大規模であると、全ての項目の組について類似度を調べるのは非常に大きなコストがかかる。そこで、類似する物のみに対して、類似度を調べるという問題設定を行い、その元で高速なアルゴリズムを開発する。これにより、ある程度一般的なデータに対して、基礎的な類似性を高速に計算できるようになることを目標とする。大規模データの解析を行いたいときには、このような基礎的な計算では不十分なことが多く、このような計算結果らより具体的、あるいは意味的な構造を見つけ出して可視化する必要がある。そこで、本研究では、そのような有効な解析手法についても研究し、幾つかの具体的なターゲットとなるデータとタスクを選び、それらに対して、類似性の計算結果を用いた解析手法の開発を行う。

### 2. 研究成果

#### (1) 概要

巨大データ解析のため、大規模データの類似性を網羅的に調べる効率良いアルゴリズムを構築することをテーマとした。大規模データの解析を行いたいときには、計算結果らより具体的、あるいは意味的な構造を見つけ出して可視化する必要がある。そこで、可視化の有効な解析手法についても研究し、幾つかの具体的なターゲットとなるデータとタスクを選び、それらに対して、類似性の計算結果を用いた解析手法の開発を行った。

本研究では、まず「文字列パターンが現れる場所の組合せ」を特定することとし、非常に短い時間で、今までは考えられなかったような長いパターンをたくさん見つけるアルゴリズムを開発することに成功した。これにより長い頻出文字列パターンのマイニングが初めて可能となった。

文字列の類似性を解析するアルゴリズムを文字列の圧縮アルゴリズムにも適用し、最適な

選択を行うアルゴリズムの開発に成功した。

グラフの列挙アルゴリズムについても研究を行った。類似するグラフの同型性の判定を大幅に削減する手法を考案し、実装を進めている。現在はより高速なプログラムを実装すべく開発を継続している段階である。

他の類似尺度については、まだ決定的な成果を得られていない。今後、他の類似性尺度、および包含や対立など他の関係性について網羅的な計算を高速で行う手法の研究を進めていく。また、実際のデータ解析についても研究を展開していく。

## (2) 詳細

類似性の解析を行う基礎的な手法については、さきがけ研究前にプロトタイプの開発が終了していた文字列の類似性を解析するアルゴリズムを一つの基盤として開発を行った。まず、一つ目の課題として、文字列のアルゴリズムを利用して、他の構造を持つデータの類似性を調べる高速アルゴリズムの開発を行った。

類似性を解析する方法として、Local Sensitive Hash(LSH)という方法が知られている。これはある種のランダムに発生した関数を用いて、データを文字列符号に変換するというもので、類似する2つの項目の符号が高い確率で(ハミング距離の意味で)短い距離を持つようになる。通常、LSH は「同一の LSH 符号を持つ項目の対」を比較することで、類似しない項目対の比較を省略する。しかし、類似する2つの項目がまったく同一の LSH 符号を持つことはやはり確率が低く、その確率を上げるために異なる LSH 符号を非常に沢山作る、ということが行われており、これが LSH を使った手法の効率を著しく下げていた。今回の研究では、「同一の LSH 符号を持つ項目の対」だけでなく、「類似する LSH 符号を持つ項目の対」を比較することで LSH 符号の精度を高め、効率性を上昇させることに成功した。従来は類似する LSH 符号を持つ項目の対を見つけることが高いコストを要求したが、以前開発した文字列の類似性を解析するアルゴリズムを用いることで、短い時間で終了することを可能とした。これにより、同じ精度で計算した場合、場合によってはメモリ効率が 10 倍、速度が 100 倍になる程度の改良を実現した。

文字列の類似性は、ゲノム配列を初めとする、中規模な類似構造を持つような文字列の距離計算にも適用した。文字列の類似性を計算する場合、編集距離と呼ばれる、2つの文字の対応を取るために置換、挿入、削除の回数をどれだけ少なくできるか、という尺度が類似度として用いられることが多い。しかしこの距離は計算に時間がかかり(長さの 2 乗)大規模な文字列データでは適用が難しい。今回は、中規模な類似構造、つまり全体の 1000 文字から全体の 10%程度の長さの類似構造をいくつか持つような文字列の類似性を調べる場合に焦点を当て、新しいアルゴリズムの開発を行った。このような類似構造を持つ場合、その類似構造は距離計算を行う際に「対応する」と見なされるべきである。ただし、複数の場所と類似する部分は、その中のどれか1つとのみ対応することになる。このような観察から、まず最初に中規模な類似構造を計算し、全体的な類似度を計算する際には、どの中規模構造をつなげるべきか、という点にのみ焦点を絞って距離計算をする、というモデルを考案した。中規模な類似構造の発見には、先の文字列の類似性を解析するアルゴリズムを用いる。これは非常に短い部分列の類似性のみを発見するので、見つかった短い類似構造を進展することで中規模の類

似構造を得ることとした。全体の計算に関しては、新しいアルゴリズムを構築し、 $O(kn \log n)$ 、あるいは  $O(n\sqrt{n})$  ( $n$  は中規模構造の数、 $k$  は類似度の中で局所的に許す異なりの数) の時間で動くようなアルゴリズムを開発した。実験的にもパフォーマンスは良好で、大規模な文字列データでも実時間で比較をすることができた。

同じような発想で、データマイニング界の難問とされる、頻出文字列マイニング問題に対しても取り組みを行った。文字列マイニングは、文字列データの中に多数現れる文字列パターンを見つける問題であり、この際、多少エラーを許す、つまり似た文字列がデータ沢山現れるような文字列パターンを全てを見つける問題となる。山登り探索といういわゆる既存手法に基づいて問題設定を行うと、短めのパターンがほぼ全て頻出パターンとなり、アルゴリズムはそれらを全て探索する必要があるために非常に遅くなる。そのため、既存手法を用いることができず、長らく未解決問題であった。

今回の研究では、文字列パターンが現れる「場所の組合せ」に注目した。パターンが現れる際に類似性を許していることから、異なるパターンがまったく同じ場所(の組合せ)に現れることがありうる。このような場合、本来パターンは一つ出力すれば十分なはずである。この観察に基づき、本研究では、まず「文字列パターンが現れる場所の組合せ」を特定することとした。具体的には、文字列データから類似する場所を全て見つけ出し、たくさんの場所と似ている場所を抽出することとし、これら「似ている場所」が作り出すグループからパターンを生成することとした。結果、非常に短い時間で、今までは考えられなかったような長いパターンをたくさん見つけるアルゴリズムを開発することに成功した。実際に見つかったパターンはデータベースに多く現れており、これにより長い頻出文字列パターンのマイニングが初めて可能となった。

文字列の類似性は、文字列の圧縮にも用いてみた。ゲノムなどのランダム性を含有する文字列データは、そのランダム性ゆえに通常のハフマン符号や繰り返し構造に基づく方法ではほとんど圧縮できない。その一方で、ゲノム配列などは、比較的大きな類似構造を持つことも分かっている。類似構造があるのであれば、似ている物が存在する部分については、「〇〇の部分にこういう変更を加えた」という形で表現することで、比較的コンパクトに圧縮することができる。文字列の類似性を解析するアルゴリズムを用いれば、このような類似性は網羅的に見つけることができるので、これを使って圧縮アルゴリズムを作ることにした。1つの部分は複数の部分と類似することがあり、また類似構造の始まり、終わりもまちまちである。その中からどの類似構造を使うと効率が良くなるか計算するため、動的計画法という手法を用いたアルゴリズムを開発し、最適な選択を行うアルゴリズムの開発に成功した。

現在最も良いとされている圧縮アルゴリズム、および過去に提案されてきたゲノム圧縮アルゴリズムは、圧縮によって 10%程度の削減しかできなかったところを、今回のアルゴリズムは 20%近い削減を可能とした。

グラフの列挙アルゴリズムについても研究を行った。グラフでは、等しいグラフは同型であると呼ばれ、同一視されるが、同型性の判定は難しい問題として知られている。そのため網羅的に列挙することもまた難しかった。似たグラフが同型であるかどうかを判定し、重複を回避するのにコストがかかるのである。今回は、類似するグラフの同型性の判定を大幅に削減する手法を考案し、それをもとに実装を進めている。予備実験により大幅な計算の省略ができることが確認できている。現在はより高速なプログラムを実装すべく開発を継続している段階で

ある。

### 3. 今後の展開

今回の研究で、文字列を中心とした類似性の解析、および類似性を用いたデータ解析手法については、いくつかのマイルストーンができたと考えている。しかしながら、他の類似尺度については、まだ決定的な成果を得られていない面は否めない。今後は他の類似性尺度、および包含や対立など他の関係性について網羅的な計算を高速で行う手法の研究を進めていく必要がある。また、解析手法がある程度整いつつあることを鑑み、実際のデータ解析についても研究を展開する必要がある。近年ビッグデータと呼ばれるビジネス・研究分野が盛んであるが、このような解析に高速アルゴリズムを必要とする分野では、アルゴリズムを基礎分類の軸としたような系統立てを持つような研究の俯瞰が必要であると考えている。ビッグデータの解析に関して、どのようなアルゴリズムがどのような課題、データに対して有効であるのかを解明し、それを元にしてモデリングを行う標準的、あるいは発展的なモデリング手法を系統立てる。また、それに伴い、各種のビッグデータ自身の性質に関しても、計算の側面、つまりどのようなアルゴリズムが有効でどのような手法がどのような精度を持つのか、という側面に関して明らかにしていく必要があると考えている。データの性質と計算手法の性質、両面からの研究を進めることで、より迅速かつ明解に、ビッグデータ解析の研究が進んで行くであろう。

### 4. 自己評価

研究開始時、当初の目標はアルゴリズムの発展であった。文字列の類似性を解析するアルゴリズムのプロトタイプがあったため、このアルゴリズムの完成度を高めること、他の類似性・関連性の計算へ拡張すること、あるいは他のアルゴリズムを考案すること、類似性の解析を用いたアプリケーション、つまり計算結果を使った新しいデータ解析手法を開発することであった。アルゴリズムの研究では、アルゴリズムの開発は単発で終わることが多く、実際に実装を作成して応用分野で使うこと、およびそのアルゴリズムを用いたアルゴリズムを開発する、という形の縦横の展開を持つものは少なかった。その意味で今回の研究は当初目的を達成し、意義ある物となったと考えている。

研究が終了して鑑みると、このようなアルゴリズムの展開の他に重要な課題が見えてきた。データ解析においては、単純・基礎的・効率の良いアルゴリズムとそれを用いた解析手法の存在が重要であると共に、それらアルゴリズムを用いたモデルの開発技術や、計算やアルゴリズムの側面から見たデータの性質(どのような手法でどのような精度がでるか)といった点も重要であることが分かってきた。現在ビッグデータの解析においては、様々な手法やデータが乱立しており、さながら戦国時代のようなものである。このような時分において、データ解析手法の研究に対して明解な方向性を得ることができる概念を得られたことは、今後の研究の展開に対して大きな成果であると考えている。

### 5. 研究総括の見解

大規模データ内の汎用な事例間類似性高速計算アルゴリズムの開発が研究課題である。

今回の研究で、文字列を中心として、非常に短い時間で、今までは考えられなかったような長いパターンをたくさん見つけるアルゴリズムを開発することに成功し、長い頻出文字列パターン



のマイニングが初めて可能となっている。また、これを文字列の圧縮アルゴリズムにも適用し、最適な選択を行うアルゴリズムの開発に成功している。

今後、他の類似性尺度などの関係性についても手法の研究の範囲を拡大してほしい。  
また、実際のデータ解析についても研究を展開してほしい。

## 6. 主な研究成果リスト

### (1)論文(原著論文)発表

1. 著者、発表論文タイトル、掲載誌名、巻号頁、発行年等 Takeaki Uno, Ryuhei Uehara, Shin-Ichi Nakano: Bounding the Number of Reduced Trees, Cographs, and Series-Parallel Graphs by Compression, WALCOM 2012, Lecture Notes in Computer Science 7157, pp. 5-16 (2012)
2. Yasuo Tabei, Takeaki Uno, Masashi Sugiyama, Koji Tsuda: Single versus Multiple Sorting in All Pairs Similarity Search, Journal of Machine Learning Research – Proceedings Track 13, pp. 145-160 (2010)
3. Takeaki Uno: Multi-sorting algorithm for finding pairs of similar short substrings from large-scale string data. Knowledge Information Systems 25(2): 229-251 (2010)
4. 松井鉄史、宇野毅明, 計算幾何学的な手法を用いた高速相同性計算手法, 情報処理学会バイオ情報学研究会 (2010)
5. Takehiro Ito, Shin-Ichi Nakano, Yoshio Okamoto, Yota Otachi, Ryuhei Uehara, Takeaki Uno, Yushi Uno: A Polynomial-Time Approximation Scheme for the Geometric Unique Coverage Problem on Unit Squares. SWAT 2012, Lecture Notes in Computer Science 7357, pp. 24-35 (2012)

### (2)特許出願

なし

### (3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

受賞: 文部科学大臣表彰 科学技術部門 若手科学者賞受賞.

巨大データ解析に対する超高速アルゴリズム構築法の研究 (2010年4月13日)

招待講演: Deep of Enumeration Algorithms, ENUMEX(ヨーロッパの研究者向けの集中講義), イタリア (2012年9月26日)

## 研究報告書

### 「映像分析による知識の抽出と、その利用による新たな映像合成」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 岡部 誠

#### 1. 研究のねらい

本研究の目的は、大量の映像データを統計的に分析して知識を抽出し、ユーザがその知識を利用することで容易な、直感的な映像製作を可能にする環境を作ることである。現在、撮影技術の進歩とインターネットの普及により、大量の映像データに容易にアクセスできる。また、テレビ局、CG プロダクションでも、独自に映像データベースを作っており、それらの加工・編集を通して、独自に新たな映像を作っている。しかし、そこで実際に行われている作業を見てみると、タグ付けされた言葉による映像の検索、見つけた映像の切り貼りと言った、技術的にはごく簡単なことしか行われていない。

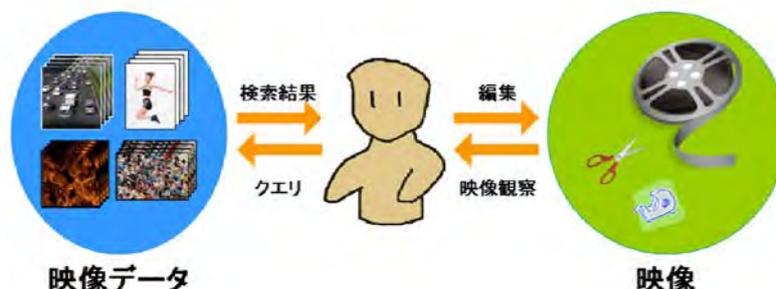


図: 現在の映像データベースの使われ方。ユーザは編集中の映像を観察し、そこに加えたい内容の映像を言葉によって検索する。システムはタグ情報に基づき検索、結果を返し、ユーザはその映像を切り貼りして編集を続ける。ユーザは、以上のループを繰り返す。

本研究では、人手によるタグ付け作業に頼らず、大量の映像データの中から抽象的な知識を自動的に抽出し、それらを効率よく利用するためのユーザインタフェースの提案を行う。今までの切り貼りを越えた、映像同士の混ぜ合わせを可能にすることで、より高度な映像製作を可能にする。映像製作技術と「大量の映像データベースの創造的な再利用」の間の橋渡しをめざす。

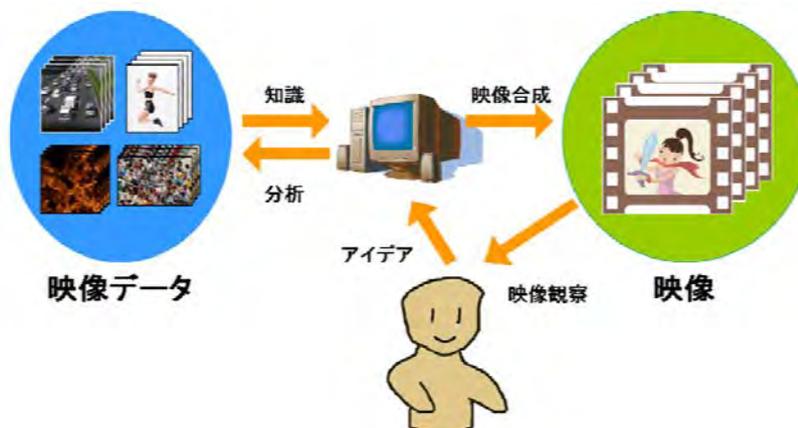


図: 提案手法による映像データベースの使われ方。コンピュータはあらかじめ、大量の映像データを分析し抽象的な知識の抽出を行う。ユーザは編集中の映像を観察し、そこに欲しい内容を、見た目や

動きの指定による、直感的なユーザインタフェースでコンピュータに指定する。コンピュータは抽出済みの知識に基づき映像合成をし、結果を更に編集する。

## 2. 研究成果

### (1) 概要

大量の映像データを統計的に分析して知識を抽出し、ユーザが直感的な映像製作を可能にする環境を作ることがテーマである。このテーマに沿って、「ビデオデータベースを用いて流体画像をアニメーションさせる手法」、「動画鑑賞のためのスケッチインタフェース」、「入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術」、「アマチュア演技者の動作を格好良く見せるように動画を編集する手法」、「勾配画像処理に基づく動画からの流体抽出」等の基本的な技術の研究開発を行った。

「ビデオデータベースを用いて流体画像をアニメーションさせる手法」は局所的な動画検索である。ユーザが絵と簡単なアノテーションという入力力でコンピュータにアイデアを与えると、コンピュータが自動的にアニメーションを制作してくれる。

「動画鑑賞のためのスケッチインタフェース」は画像検索と軌跡検索の組み合わせで、新しい映像閲覧のためのユーザインタフェースの研究である。ユーザはスケッチというシンプルなアノテーションを与えるだけで、予め動画を解析したコンピュータがユーザの見た動画の即座に検索してくる事ができる。

「入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術」は動画の編集技術の提案である。映像が音楽の楽しみ方に与える影響について、引き続き分析と実験を行っている。

「アマチュア演技者の動作を格好良く見せるように動画を編集する手法」は、アマチュア動画とプロ動画の両方の動作を分析して抽出し、前者を後者に合わせて編集する技術である。

「勾配画像処理に基づく動画からの流体抽出」は爆発や煙、炎等の半透明領域が広範囲に渡る動画において、その流体を抽出する技術である。リッチな動画データベースを構築するのに助ける技術である。

2次元動画データベースの解析と合成の研究に従事したと共に、2次元動画の限界も見えている。今後、上記の2次元流体動画の応用研究を、3次元流体動画に拡張していく。

### (2) 詳細

① . ビデオデータベースを用いて流体画像をアニメーションさせる手法

「論文 1, 学会発表 1, 11, 受賞 1, 2」

下図(a)はキャンプファイアの写真, (b)は 400 年くらい前に描かれた油絵である. 1枚の絵であれば, (b)の油絵のように, アーティストは独自のタッチで詳細に描き込むことができる. しかし, そのようなシーンのアニメーションを作ろうとすると, 流体シミュレーションを走らせて3次元 CG で再現するのも難しいし, かと言って何枚も同じような絵を描いてパラパラアニメ化するのもとても大変である. そこで, 1枚の絵の持つ雰囲気, 質感をなるべく損なわないようにアニメーションを作るため, 我々はビデオ検索技術と画像合成技術を組み合わせて技術を開発した.



図. (a)キャンプファイアの写真. (b)川の急流を描いた油絵.

システムは下図(a)に示すように, 流体ビデオのデータベースを持っている. ユーザが指定するのは3つで, (b)の流体画像とその上の大体の流れ方向を指定するオレンジの矢印, (c)の流体領域である. 後は自動的にシステムが(d)の流体アニメーションを作る.

<http://130.153.152.4/~okabe/FluidAnimeVideoDB/index-j.htm> にアクセス頂くか、「岡部 ビデオデータベース」で検索するとデモ動画が見られるのでご覧頂きたい.

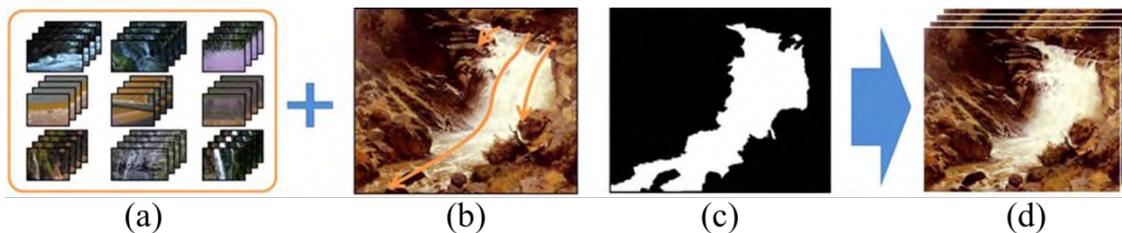


図. システム概要.

この技術の肝は, 局所的な動画検索である. 下図は, 3つのビデオデータを用いて小川の写真をアニメーションさせようとしている. この3つの小川のビデオは写真の小川とは全く異なるものだ. しかし, 局所的に見ると似ている部分もある. 例えば, 黄色, オレンジ, 青の丸印で囲んだ領域では, それぞれビデオと写真の間で水しぶきの調子が似ている. このような類似性にもとづいて, 各丸で囲んだ部分のビデオを写真上へコピーする, というのがアイデアである. 写真上の全ての場所へ, いろんなビデオから部分的なコピーをしつつ, 画像合成技術を使って切れ目が見えないように合成することで最終的なアニメーションを得ることができる.

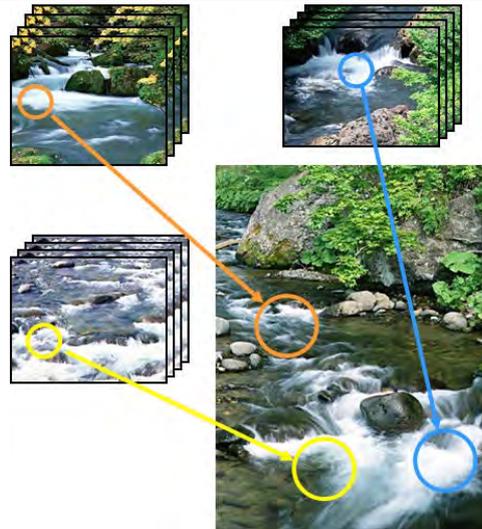


図. ビデオの部分的なコピーと画像合成によって小川のアニメーションを作る.

## ② 動画鑑賞のためのスケッチインターフェース

「論文 2, 特許 1, 学会発表 2, 9」

今, 下図(a)のように, 車が左から右へ走るシーンを見ているとする. この時「この車が手前へ走ってくる, (b)みたいなシーンがあったなあ, あれを見たいなあ」と思うことがある. 普通だと, 早送り, 巻戻しボタンや再生バーを操作してシーンを探す. それでも見つからない場合は YouTube へ行って検索するかも知れない. しかし, ここではもっと迅速に目的のシーンへ移動する手法を提案する.

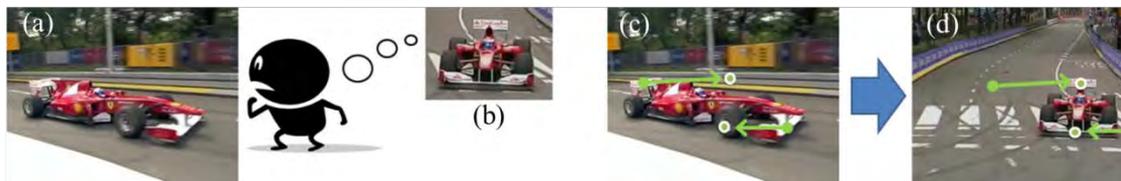


図. (a)左から右へ走る車を鑑賞中に, (b)のような手前に走ってくる車のシーンを見たくなった. (c)前後のウイングをつまんでクルッと回すイメージで緑の矢印を描くと, (d)の目的のシーンが検索される.

ユーザの見たい(b)のシーンとはつまり, 今見ている(a)の車を, クルッと回して前を向かせたようなシーンのことである. そこで, このクルッと回す感じを(c)のように緑の矢印でもって, ユーザにスケッチしてもらおう. 「回転して, 前後のウイングが丁度この辺に来ると, 私の見たいシーンになるなあ」という感じである. すると, システムが即座に目的のシーンを検索してくれる. デモ動画は「川手 岡部 尾内」で検索すると川手氏のホームページが出てくるので, そちらでご覧頂きたい.

この技術の肝は, 画像検索と軌跡検索を組み合わせたところにある. 画像検索だけだと, (a)の画像が与えられても (b)の車を直接検索することは難しい. この様子を下図に示す. 画像検索でよく用いられる, 局所的な画像特徴のマッチングを試みた例である. (a)に示す車は左右反転

しているものの、車自体は似たような姿勢をしていてマッチングがたくさん採れるので検索することができる。一方、今ユーザが見たいと思っている(b)は、局所的なマッチングは3つしか採れなかったのが通常は検索結果とならない。

しかし、この(a)と(b)の関係に注目する。実は (b)の車が手前に走ってきて左側へカーブしていくのが(a)、つまり(a)と(b)は一続きのシーンなのだ。そこで我々のシステムは、まず局所マッチングに基づいて(a)を検索し、次に(a)を起点に早送り、巻戻しを自動で行いながら「ユーザがスケッチした緑の矢印みたいな軌跡を描いて前後のウイングが動いたようなシーンはないかしら？」とチェックする。すると、(a)を少し巻戻したところに(b)が見つかるというわけである。

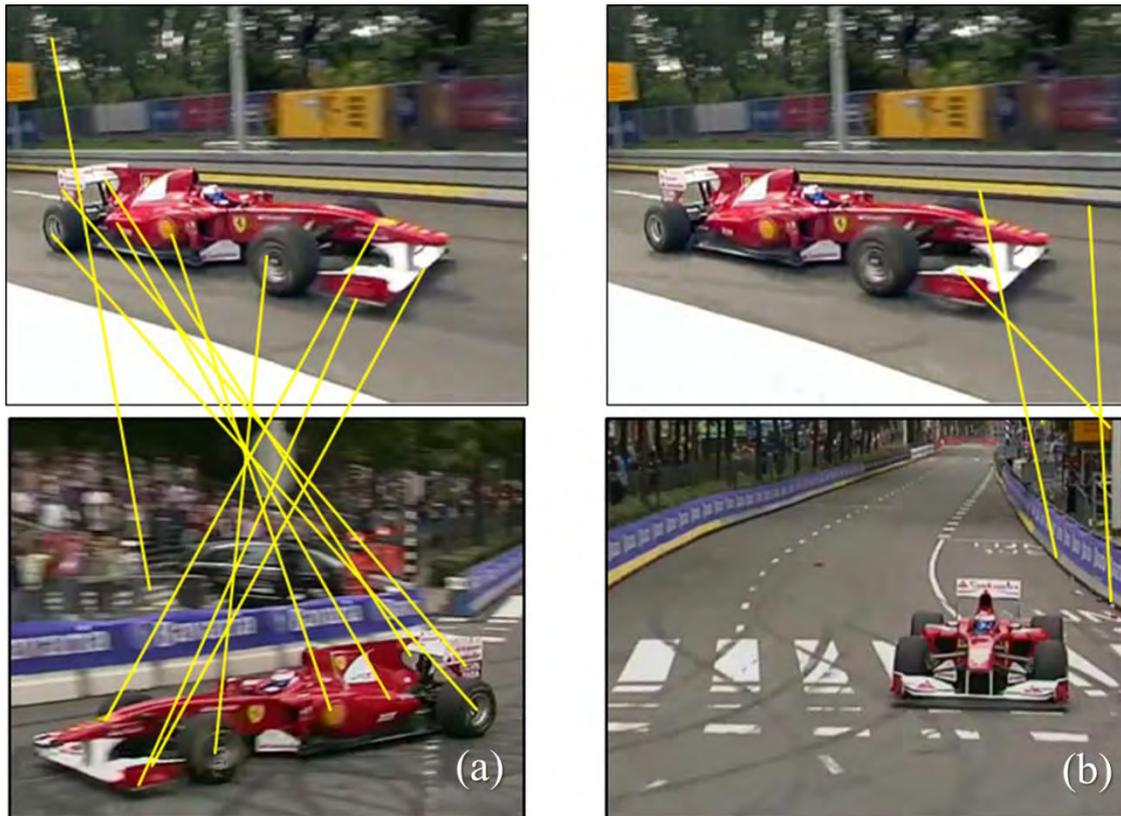


図. 画像検索で用いられる局所的な画像特徴のマッチング.

### ③ 入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術

「論文 5, 学会発表 3, 5, 8, 12, 受賞 4」

入力音楽に合わせて、その音楽の演奏動画を半自動的に生成し、演奏を耳と目の両方で楽しむことができる。この手法ではピアノ、ギター、ベース、ドラムなどの演奏動画を収集して動画データベースを作っている。下図の Database に示すように、動画のオーディオトラックを予め分析することで、図中の(b)などの音符の発音タイミングを抽出しておく。この発音タイミングは実際に演奏者が演奏しているであろう楽譜上の音符の位置に相当するようなものだ。今、下図の Input Music に示すような入力音楽が与えられたとすると、システムは同様のアルゴリズムで図中の(a)に示すように発音タイミングを抽出する。この発音タイミングを特徴量として動画データベースを検索することで、類似した発音タイミングを持つ動画を検索して下図中の「??」部

分に割り当てる。ここでは(a)と(b)は類似しているのので、(b)の動画を割り当てれば、(a)を演奏しているように見えるに違いない、というわけである。この作業を入力音楽の全てのパートに繰り返すことで、演奏動画を生成する。デモ動画は「yamamoto okabe onai」などと検索すると山本氏のホームページが出てくるので、そちらでご覧頂きたい。

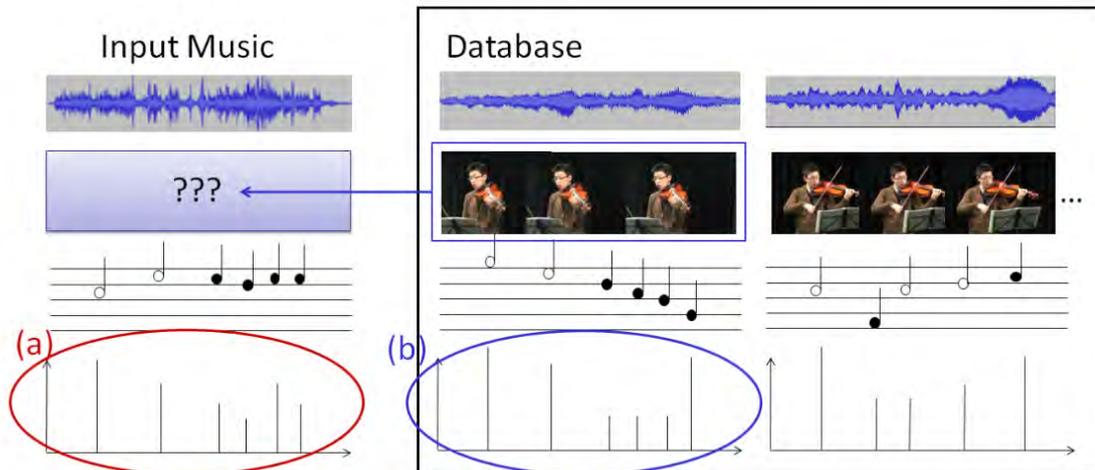
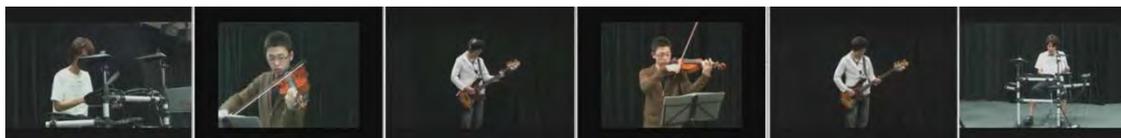


図. 入力音楽とデータベース. 両者において、音符の発音タイミングを自動的に抽出して検索を行う。

下図に合成動画のフレームをいくつか示す。音楽をただ聴くだけだと、それほど注意して聴いていなかった低いベースの音も、映像中のベース奏者の指の動きと一緒に聴くとより鮮明に聴こえてくる。映像（偽モノですが）が音楽の楽しみ方に与える影響について、引き続き分析と実験を行っている。



下図. ここでは、ビートルズの「レットイットビー」を入力音楽としてバンド演奏を合成している。

#### ④ アマチュア演技者の動作を格好良く見せるように動画を編集する手法

「論文 4, 学会発表 4, 7」

プロの演技者はダンサーでもアクションスターでも非常に格好良い。体は柔らかく筋力があり動きにキレがある。ところが、我々がそれに憧れて真似ようとしても上手くいかない。体は硬く筋力はなく動きにキレが無いのである。そこで、コンピュータに手伝ってもらい、せめて動画の中でだけでも格好良くなろう、という願いに基づいた研究である。

下図(a)では、アマチュアが回し蹴りに挑戦している。しかし、(b)のプロのように足は高く上がらないし、動きも遅くて格好悪い。そこで、このプロの動作を抽出してアマチュアの動画に適用することで編集したものが(c)である。これで、足は高く上がっているし、動きも速く格好良くなった。「mizui okabe onai」で検索すると水井氏のホームページがあるので、是非デモ動画をチェッ

くして頂きたい。



図. プロの動作を抽出してアマチュアの動画に適用することでアマチュアを格好良くする.

技術的には、アマチュア動画とプロ動画の両方の動作を分析して抽出(下図(a)と(b))して、前者を後者に合わせて編集する(下図(c)). 動作の分析は図のように、人物の動きをスケルトンで追跡することで行う。この際、全自動でコンピュータが動きを完璧に追跡することは難しいので、間違いが起こった際はユーザがマウスでスケルトンを編集できるようなツールを作っている。スケルトンが抽出できたら、(a)のスケルトンが(b)のスケルトンに重なるように(a)の画像を変形する。すると(c)ができ、足が高く上がったように見える(背景もそれに合わせて歪んでいるが、人物のみを自動で切り出して最終結果を合成する)。

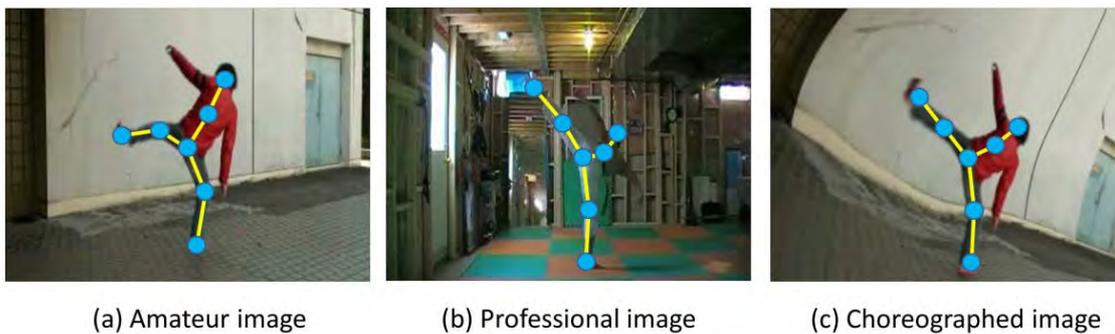


図. スケルトンを用いた動作の追跡と変形への適用.

#### ⑤ . 勾配画像処理に基づく動画からの流体抽出

「論文 3, 学会発表 1, 6」

動画から爆発や煙等の半透明部分が広範囲に現れるような流体を抽出する手法を提案した。例えば下図では、ビル前で起こった爆発の映像から、ビルを消し、爆発のみを抽出できれば、後でこの爆発を動画制作に利用できる。しかし、流体の背後にはこのビルのような複雑な背景がうっすらと見えるのだが、これと流体を分離するのは時間と手間の掛かる作業となる。

我々は全自動でこの問題を解決する手法を提案した。ここで動画を撮影したカメラは静止しているものとし、我々はまず背景画像を推定し、次に流体の前景とアルファマトを推定する。本手法では、動きの少ないピクセルで背景が見えることを用いて背景を推定する。次に推定された背景との差分としてアルファマトを推定する。この差分によるアルファマトには背景模様が出現してしまう。そこでこの模様を消すために勾配画像処理を用いる。勾配画像中で背景模様に関する勾配を弱め、その上で積分することで精製されたアルファマトを得る。前景画像も同様の勾配画像処理を用いて得られる。

本手法を用いて爆発、煙、炎等の半透明領域が広範囲に渡る動画において、その流体の抽出

ができることを示す。また、下図に見えるように、既存の Trimap を用いた手法と比較しても、提案手法がクオリティの高い流体抽出ができることを示した。

<http://makotookabe.com/FluidExtraction/index.htm> にて動画及びソフトウェアを公開中。

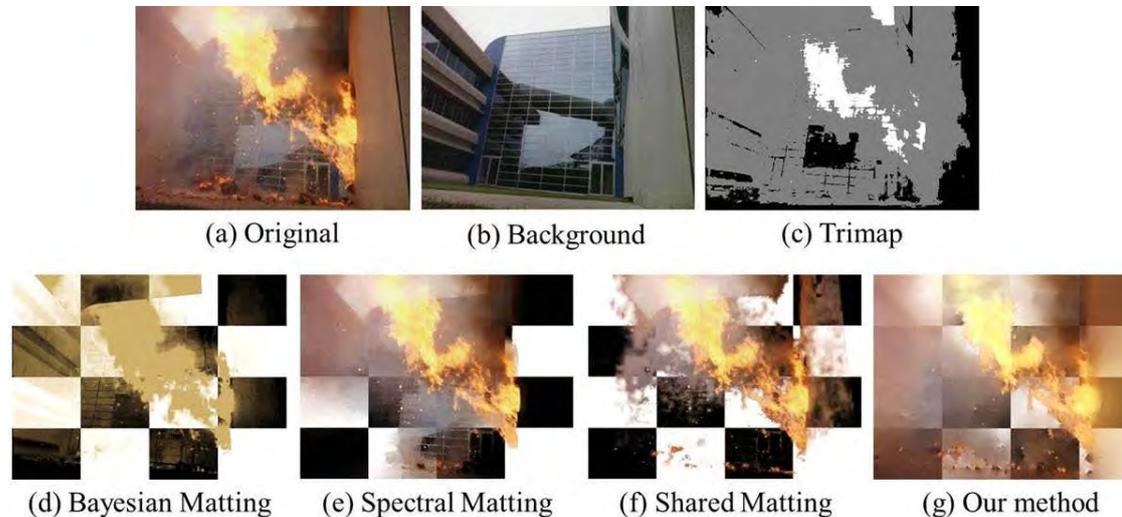
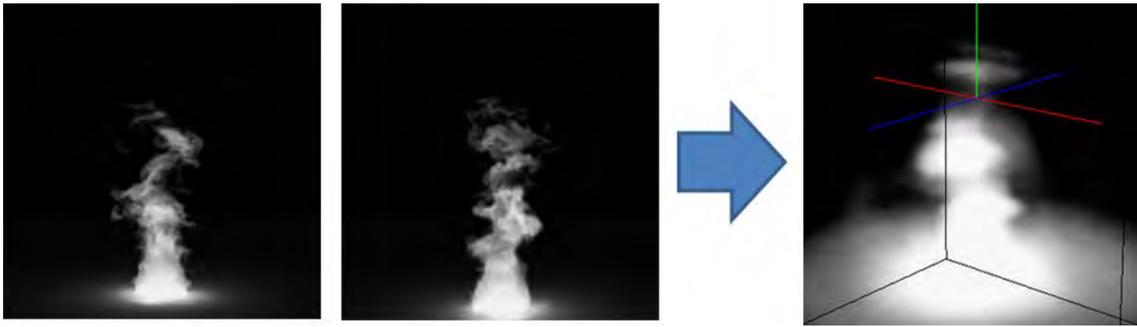


図. 既存手法との比較。オリジナルフレーム(a)と背景画像(b)との差分からトライマップ(c)を作成して既存手法への入力とした。Bayesian Matting で作成した合成画像(d)。Spectral Matting で作成した合成画像(e)。Shared Matting で作成した合成画像(f)。本手法で作成した合成画像(g)。

### 3. 今後の展開

本年度までの3年半、2次元動画データベースの解析と合成の研究に従事したと共に、世界中で映像データベースに基づく映像合成が盛んに研究されている事実も踏まえ、動画データベースの研究を続けることが未来の映像表現に貢献できると確信しつつ、同時に2次元動画の限界も悟るに至った。申請者の考える2次元動画の限界は大きく2つ。1つは上述した「後で視点を変えられない」という3次元的表现への限界である。もう1つは、2次元動画を解析・学習させても、得られるコンピュータの映像理解力に限界があるという事である。我々人間も両眼立体視や焦点調節を通じ、周囲の環境を3次元的に学習して、その結果、現在の視覚能力を得ているのである。近年、我々の研究環境もコンピュータに人間と同様、3次元環境を学習させるのに都合の良い世界になってきた。最も顕著な事は3次元撮像技術の発展である。近年ハリウッドで製作される映画の多くが3次元映画で、全ての映像が2台のカメラでステレオ撮影された3次元映像であることが普通である。Microsoft Kinect に代表される3次元スキャナも安価に普及している。この情勢を見れば、今後、数十年の間に3次元動画コンテンツが普及していくのは明らかであり、現在の2次元画像・動画の世界は大きく変わっていくと思われる。

今後の大きな予定の1つは、上で述べた2次元流体動画の応用研究を、3次元流体動画に拡張していくことである。3次元流体を測定する技術はいくつか提案されているが、特別な機材を必要とするなど、扱いやすいものがない。したがって、安価な測定技術自体を作るという研究が可能である。一方、さまざまな流体を測定 & 収集して3次元動画データベースを作り、検索、加工、動画合成のためのオーサリングツールを作る等の研究を進めることも可能である。現在、2つのカメラで撮影した動画像から煙の3次元化を行うところまで行っている(下図)。



#### 4. 自己評価

「①ビデオデータベースを用いて流体画像をアニメーションさせる手法」は「研究のねらい」に書いた内容を、ほぼそのまま実現しようとした研究である。ユーザは絵と簡単なアノテーション、という入力でコンピュータにアイデアを与える。後は流体動画データをあらかじめ分析したコンピュータが自動的にアニメーションを制作してくれる。ユーザが結果を修正したければ、アノテーションの修正及びパラメータ調整をして、コンピュータに再計算を依頼できる。研究結果は雑誌、及び、2つの学会発表、2つの表彰という形でまとめる事が出来た。プレゼンをする時(フレンドリーな会場だと)「おー」と歓声上がる事もあったので、良い研究が出来たのではないかと考えている。今後は手法として誰しも簡単に使えるツールという形で公開できるように、手法を簡略化したいと考えている。

「②動画鑑賞のためのスケッチインタフェース」は新しい映像閲覧のためのユーザインタフェースの研究である。こちら、ユーザはスケッチというシンプルなアノテーションを与えるだけで、予め動画を解析したコンピュータがユーザの見た動画の即座に検索してくる事ができる。物体の回転等を考慮した検索が可能なのだが、3次元形状を分析する必要が無く、2次元的分析のみで済むという辺りがウリである。この研究も国際学会での発表と国内特許をとる形で成果にまとめた。現在より大規模なデータで実験を行ったものの論文を執筆中である。

「③入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術」「④アマチュア演技者の動作を格好良く見せるように動画を編集する手法」も今までにない動画の編集技術の提案を行い、国際会議での発表を行う事が出来た。特に④の研究では学会会場の発表等で、聴衆から笑い声を聞く事ができ、本来の研究の目的を達成できたと考えている。「⑤勾配画像処理に基づく動画からの流体抽出」は上記のような研究を進めるために、リッチな動画データベースを構築するのを助ける手法であり、本研究のテーマの補助的な研究であるが、今まで解けなかった問題を解く事ができた。こちらは学会発表を行い、ソフトウェアを公開している。

#### 5. 研究総括の見解

動画 DB を参照しながら静止画を動画として合成する研究である。

大量の映像データを統計的に分析して知識を抽出し、ユーザが直感的な映像製作を可能にする環境を作ることがテーマである。このテーマに沿って、「ビデオデータベースを用いて流体画像をアニメーションさせる手法」、「動画鑑賞のためのスケッチインタフェース」、「入力音楽をあた

かもバンド演奏しているかのように見える動画の生成技術」、「アマチュア演技者の動作を格好良く見せるように動画を編集する手法」、「勾配画像処理に基づく動画からの流体抽出」等の基本的な技術の研究開発を行った。それぞれ、「目で見て分かる」成果をあげている。

今後、このさきがけ研究を通じて得られた、2次元動画の解析・学習の限界とコンピュータに人間と同様に3次元環境を学習させるべきとの知見に基づき、2次元流体動画の応用研究を3次元流体動画に拡張していくと同時に、これらの成果を統合して、どのような方向に向かうのかを明確にしながらか研究を進めてほしい。

## 6. 主な研究成果リスト

### (1) 論文(原著論文)発表

1. Makoto Okabe, Ken Anjyo, Rikio Onai, "Creating Fluid Animation from a Single Image using Video Database", Computer Graphics Forum, vol. 30, no. 7, pp. 1973-1982.
2. Makoto Okabe, Yuta Kawate, Ken Anjyo, Rikio Onai, "Video Retrieval based on User-Specified Appearance and Application to Animation Synthesis", In Proc. of MMM 2013.
3. Makoto Okabe, Ken Anjyo, Rikio Onai, "Extracting Fluid from a Video for Efficient Post-Production", In Proc. of DigiPro 2012.
4. Kenta Mizui, Makoto Okabe, Rikio Onai, "Choreographing Amateur Performers Based on Motion Transfer between Videos", In Proc. of MMM 2013.
5. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of Performers Appearing to Play User-specified Band Music", ACM SIGGRAPH 2012 poster.

### (2) 特許出願

研究期間累積件数: 3 件

1.

発 明 者: 平野 廣美, 三條 正裕, 川手 裕太, 岡部 誠, 尾内 理紀夫  
 発明の名称: 動画検索装置、動画検索方法、記録媒体、ならびに、プログラム  
 出 願 人: 楽天株式会社  
 出 願 日: 2012/1/20

2.

発 明 者: 平野 廣美, 岡部 誠  
 発明の名称: 画像処理装置、画像処理方法、及びプログラム  
 出 願 人: 楽天株式会社  
 出 願 日: 2012/8/24

3

発 明 者: 尾内 理紀夫, 岡部 誠, 栗原 竜矢, 三條 正裕, 平野 廣美  
 発明の名称: 動画処理装置、動画処理方法、ならびに、情報記録媒体  
 出 願 人: 楽天株式会社  
 出 願 日: 2012/9/7



(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

●学会発表

1. Makoto Okabe, Ken Anjyo, Rikio Onai, "Creating Fluid Animation from a Single Image using Video Database", Pacific Graphics 2011.
2. Yuta Kawate, Makoto Okabe, Rikio Onai, "Video Retrieval based on User-Specified Deformation", ACM SIGGRAPH 2012 poster.
3. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of Performers Appearing to Play User-specified Band Music", ACM SIGGRAPH 2012 poster.
4. Kenta Mizui, Makoto Okabe, Rikio Onai, "Choreographing Amateur Performers using Video Examples", To appear at SIGGRAPH Asia 2011 poster, pp. 8:1--8:2.
5. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of a Performer Appearing to Play User-specified Music", ACM SIGGRAPH 2011 poster.
6. 岡部 誠, 安生 健一, 尾内 理紀夫, "勾配画像処理に基づく動画中の流体部分抽出", ビジュアルコンピューティング '12, 2012年6月.
7. 水井 健太, 岡部 誠, 尾内 理紀夫, "動画間の動作転送に基づくアマチュア動画編集システム", ビジュアルコンピューティング '12, 2012年6月.
8. 山本 知大, 岡部 誠, 尾内 理紀夫, "バンド演奏における疑似演奏動画の合成手法", ビジュアルコンピューティング '12, 2012年6月.
9. 川手 裕太, 岡部 誠, 尾内 理紀夫, 平野廣美, 三條正裕, "静止画内オブジェクトへの指示による動画検索手法の提案", DEIM 2012, 2012年3月.
10. 鈴木 喜也, 岡部 誠, 尾内 理紀夫, "無音動画に対する効果音貼付けシステムの試作", DEIM 2012, 2012年3月.
11. 岡部 誠, 安生 健一, 尾内 理紀夫, "ビデオデータベースを用いた流体画像に基づくアニメーション生成", ビジュアルコンピューティング '11, 2011年6月.
12. 山本 知大, 岡部 誠, 尾内 理紀夫, "楽曲を演奏しているように見える動画の生成手法", ビジュアルコンピューティング '11, 2011年6月.

●受賞

1. 岡部 誠, 安生 健一, 尾内 理紀夫, "ビデオデータベースを用いた流体画像に基づくアニメーション生成", 2011年6月.(VC賞)
2. 岡部 誠, 安生 健一, 尾内 理紀夫, "ビデオデータベースを用いた流体画像に基づくアニメーション生成", 2012年6月.(グラフィクスとCAD研究会優秀研究発表賞)
3. 鈴木 喜也, 岡部 誠, 尾内 理紀夫, "無音動画に対する効果音貼付けシステムの試作", DEIM 2012, 2012年3月.(学生奨励賞)
4. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of a Performer Appearing to Play User-specified Music", ACM SIGGRAPH 2011 poster. (Student Research Competition Semi-Finalist)
5. 湊 匡平, 岡部 誠, 尾内 理紀夫, "俳句に適合する画像を合成するシステムの開発", 第3回楽天研究開発シンポジウム, 2010年12月.(優秀ポスター賞)

● 著作物

1. Computer Graphics Gems JP 2012 -コンピュータグラフィックス技術の最前線(共著)

# 研究報告書

## 「ベイジアンテレビ:取材・配信・編集を自動化した緊急情報メディア」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 北本 朝展

### 1. 研究のねらい

情報発信源の爆発的な増大に伴って、インターネットは世界の動きをリアルタイムに伝えるメディアへと発展しつつある。しかし情報発信源が大幅に増加した結果、たとえ高性能な検索エンジンが存在しても、利用者が情報を引き出す(プル)積極的な行為をしなければ重要な情報にたどりつけない問題が生じている。特にクライシス時においては、短い時間で

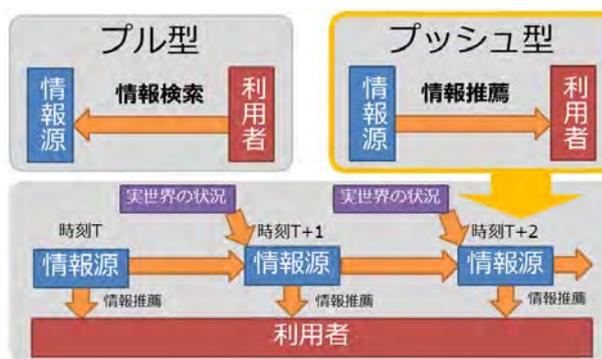


図1 ベイジアンテレビの概念図

知らない分野に適切なキーワードを設定して効率的に検索せよ、というのは現実的ではない。

そこで本研究では、特にリアルタイム性が重要となるクライシス時の情報爆発を念頭に、重要な情報が利用者に届くようなプッシュ型メディアの構築を目標とし、その概念モデルを取材・配信・編集を自動化した「ベイジアンテレビ」として提案する(図 1)。具体的な研究課題は、実世界の情報を取材・編集するエージェントの構築、データを実世界のエンティティと関連付ける解析ソフトウェアの開発、実世界の情報を取り入れつつ個別の利用者に情報を推薦するデータ同化アルゴリズム、プッシュ的・テレビ的なコンセプトに基づくデータ可視化インタフェースのデザインなどである。研究の対象とするデータは、当初は台風関連データを中心とする予定だったが、研究開始後に東日本大震災が発生したため急きょ震災関連データも研究対象に含め、事態の推移にリアルタイムで対応しながら研究課題に関する実践的な考察を深めることを狙った。

### 2. 研究成果

#### (1) 概要

利用者が情報を引き出す行為がなくても重要な情報が利用者に届くようなプッシュ型メディアの構築を目標とし、取材・配信・編集を自動化した「ベイジアンテレビ」の開発を研究テーマとした。実世界の情報を取材・編集するエージェントの構築、データを実世界のエンティティと関連付ける解析ソフトウェアの開発、実世界の情報を取り入れつつ個別の利用者に情報を推薦するデータ同化アルゴリズム、プッシュ的・テレビ的なコンセプトに基づくデータ可視化インタフェースのデザインを具体的な開発項目とし、地名情報処理のためのソフトウェア GeoNLP の開発、ソーシャルメディアを用いた気象現況把握「ふってきったー」の構築、災害関連データの取材・編集システムの構築、プッシュ化・テレビ化インタフェース「311 メモリーズ」と「311TV」の開発・公開を果たした。

地名情報処理のためのソフトウェア GeoNLP の開発では、地名解決の方が GPS 情報付きの場合よりもマッピング可能性はるかに高いことが分かった。また、災害関連データの取材・編集システムの構築では、東日本大震災後に重要となった電力データや放射線データなど、各種のオープンデータにも広がった。これらは、研究開始当初には対象としていなかったものである。これらは「311 メモリーズ」として公開し、第 16 回文化庁メディア芸術祭のアート部門審査委員会推薦作品に選ばれた。

ただし入力と出力をつなぐ部分、具体的には情報源と利用者をつなぐ「情報の交換機」となる部分は、まだ課題が多く残っている。

## 2) 詳細

### 研究テーマ1 地名情報処理のためのオープンソースソフトウェア GeoNLP の開発

実世界のイベントを報じるテキスト情報を位置と関連付けるジオタギング処理は、特にクライシス時にように迅速な状況認識が必要な場合に高い価値がある。こうした地名情報処理のためのオープンソースソフトウェア GeoNLP を構築し、近日中に公開する準備を進めている。例えば図 2 は東日本大震災関連ニュースから地名を抽出してマッピングした結果であり、円の色がその地名におけるニュース記事数を示す。



図2 GeoNLP による地名抽出

このソフトウェアが解決する問題は、自然言語文からの固有名抽出と固有名解決の 2 つである。前半の抽出では、地名辞書に含まれる地名とマッチングする単語を地名候補として抽出する。後半の解決では、複数の地名候補の中から文脈に最も適合する候補を決定する。また地名情報処理の精度は地名辞書の整備状況にも依存するため、地名辞書の共有も含めた情報プラットフォームを整備し、地名情報処理のハードルを下げることを狙う。

GeoNLP が高速な地名解決処理を実行できることを示すために、東日本大震災ビッグデータワークショップで提供された東日本大震災後 1 週間の日本語ツイート約 1 億 8000 万件に含まれる地名を解決したところ、1 時間に 1 コアあたり約 4 万ツイートを処理でき、112 コアを用いて 40 時間で全ツイートに対する地名解決処理を完了した。また、この解析結果では地名を含むツイートは全体の約 12%であった。この数字は GPS 情報付きツイートの割合 0.15%よりもはるかに大きいため、地名解決の方がマッピング可能なツイートは大幅に多いことがわかった。ただし GeoNLP は recall の向上を主に狙っているため、precision は満足できる水準には達しておらず、予備的な評価では precision は 60%程度にとどまっている。このような認識誤りを考慮すれば、地名を含むツイートは実際には 10%弱であり、12%という数字は過剰抽出によるものと考えられる。

## 研究テーマ2 ソーシャルメディアを用いた気象現況把握「ふってきったー」の構築

地名情報処理ソフトウェア GeoNLP の応用として、ソーシャルメディアから準リアルタイムで気象現況を把握するためのシステム「ふってきったー」を構築した。これはツイッターから「雨」や「雪」等の気象に関するキーワードを含むツイートを収集し、ツイートに含まれる地名を GeoNLP で解決し、その結果を地名ごとに集約してマッピングすることで、各地の雨や雪の状況を準リアルタイムで把握するシステムである。その結果、関東圏のようにソーシャルメディア人口が多い地域では、気象現象の準リアルタイム把握にソーシャルメディアが補完的に役立つ可能性があることを示した。例えば、雨と雪のツイート数を比較した図3は、どの地域が雨か雪なのかという既存のセンサ網では把握しづらい現況が、ツイート数の大小として観測できることを示唆している。このグラフから東京は雪、横浜は雨と考えられるが、実際の天候もそうであったことを気象庁官署の観測データから確認できた。その他の気象現象キーワードを用いた結果についても、「ふってきったー」では10分ごとに準リアルタイムで更新した結果を閲覧できる。

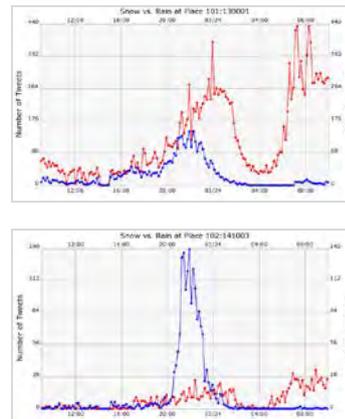


図3 ツイッターによる気象現況把握。上段は東京、下段は横浜。赤線は雪、青線は雨ツイート数を示す。

## 研究テーマ3 災害関連データの取材・編集システムの構築

台風関連データや東日本大震災関連データをネット上から取得（取材）し、そこから情報を抽出して編集する各種のシステムを構築した。その取材対象は、当初から想定していたアメダス等の気象観測データや台風関連ニュース記事等のテキストデータだけではなく、東日本大震災後に重要となった電力データや放射線データなど、各種のオープンデータにも広がった。これらのデータは現状では個別対応が必要な部分が多いが、共通部分をフレームワーク的に構築すると同時に、解析・編集の基礎データとなる形態素解析辞書の整備や地名辞書の整備などの地道な作業も推進した。

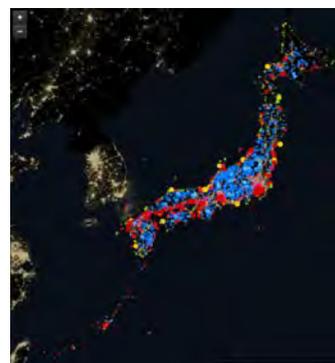


図4 電力データの可視化

またインターネット上の情報を機動的に取材するための能動的なウェブクローラーも開発した。平常時に利用するクローラーと異なる点は、気象警報等の外部からの情報に基づいてクローラーの収集範囲や頻度を動的に変更するための制御機能を備えている点にある。システムはウェブクローラー Apache Nutch を改良して構築し、クローラーとの入出力にはドキュメント志向データベース MongoDB を用いた。緊急時はデータ更新が頻繁になるだけでなく、混乱の中でデータが後から検索不能になることも多い。長期的にデータをきちんと解析するためにも、クローラーの高度化による機動的な取材とアーカイブ機能を強化していく必要があると考えている。

#### 研究テーマ4 プッシュ化・テレビ化インタフェース「311メモリーズ」と「311TV」のデザイン

本研究を始めた最大の動機は、情報のプッシュ化とテレビ化という目標にある。プル型メディアは必要な情報を検索する時間がある平常時には有用だが、情報を検索する余裕がない緊急時には、テレビのように立ち上げるだけで必要な情報が流入してくるプッシュ型メディアが必要である、というのが本研究の重要な仮説である。このような方針に基づき構築したプロトタイプが、「311メモリーズ」と「311TV」である。



図5 311TVのインタフェース

まず「311メモリーズ」は「静かに動く年表」というコンセプトを設定した。東日本大震災以後の震災関連オンラインニュース(Yahoo! News)を対象とし、自然言語処理を活用して日ごとの重要キーワードを抽出し、それを動く時間軸の上に表示した。これをプッシュ化・テレビ化インタフェースと呼ぶ理由は以下の2点である。第一に、利用者は能動的な行動を起こさなくても情報を眺めることができる。第二に、視覚的な情報と音楽という聴覚的な情報を合わせた視聴覚体験を味わえる。本システムを利用した結果から、能動的な情報探索体験よりも受動的な視聴覚体験の方が情報に対する思索を深めやすいのではないかと、という新たな仮説を得た。能動的モードでは情報の中身の精査よりも情報の探索の最適化に意識が向くため、選択肢を減らした受動的モードの方が情報について深く考えられることが原因ではないかと推測している。なお本作品は、第16回文化庁メディア芸術祭のアート部門審査委員会推薦作品に選ばれた。

次に「311TV」では、スクリーンを仮想的な地域チャンネルに見立て、スクリーン表示範囲の地域情報がスクリーンにプッシュされ可視化されるという、テレビをメタファーとするインタフェースを構築した。スクリーン上ではテキストデータは自動的にスクロールし、ティッカーエリアには緊急情報が横に流れ、ラスタデータは地図上に表示される。利用者の能動的な行動は、地図や時間を移動するなど最小限で構わない。「311TV」は利用するデータのライセンスの都合で一般には公開できないが、これを発展させた「デジタル台風TV」を近日中に公開する予定である。

またこの開発と並行する形で最近ではオープンガバメントの流れが加速しており、2012年12月からは気象庁が防災気象情報のプッシュ型試験配信を開始した。こうした公的サービスとも連携することで、重要な情報を即時的に届けるインタフェースの実現性も高まりつつある。その有効性を実際の台風情報で確かめることが今後の課題である。

### 3. 今後の展開

本研究ではベイジアンテレビの概念モデルの中でも、入力と出力の両側から研究を進めていったため、両側から成果が生まれてきた。研究開始当初「ベイジアンテレビ」は構想だけのものだったが、各種のコンポーネントが徐々につながって全体像が見え始めてきた感がある。ただし入力と出力をつなぐ部分、具体的には情報源と利用者をつなぐ「情報の交換機」となる部分は、まだ課題が多く残っている。この部分については、外部の情報を取り入れ、内部の状態を更新し、

その時点で個別の利用者が必要とする情報を推薦するアルゴリズムを、データ同化などの観点からさらに研究していく必要がある。またこの課題は、近年の潮流であるデータジャーナリズムや計算論的ジャーナリズム等のジャーナリズムの変革といった文脈から捉えることもできる。メディアとしては「テレビ」は最大にして最強の旧来メディアであり、だからこそIT業界のプレイヤーもテレビを最終目標とした計画を練っている。こうした流れも見つつ、新しいメディアを創生するための研究課題を明確化して研究を進めていきたいと考えている。

#### 4. 自己評価

本研究は、偶然ではあったが時宜を得たものであった。当初は台風関連データを対象としてクライシス・メディアを研究することを目標としたが、研究開始から1年半後の2011年3月に東日本大震災が発生し、本研究を活かすべき現場が突如として目前に出現した。この現場への対応は研究上も重要であると考え、東日本大震災関連データを対象に加えて事態の推移にリアルタイムで対応する取り組みを開始した。事前予測不可能な状況への対応を通して、実感をもって各種のメディアの限界を認識できたのは貴重な体験だった。また各種の貴重なデータを収集し、データを活用したサービスも開発し、その一つがメディアアートとして文化庁メディア芸術祭で受賞できたことは、研究対象を拡大した判断が適切だったことを示すと評価している。また GeoNLP を中心とするソフトウェア開発でも、震災時に使えなかったこと、そして大規模震災データを対象に評価を進めたことが、研究を進める上での強い動機として作用した。当初の研究課題の中に未達成部分が残ってしまったことは心残りである。しかし、この超大規模震災の期間にさきがけ研究を進めていたからこそ可能になった研究も多く、その巡り合わせには感謝の念を持っている。

#### 5. 研究総括の見解

利用者が情報を引き出す行為がなくても重要な情報が利用者へ届くようなプッシュ型メディアの構築を目標とし、取材・配信・編集を自動化した「ページアンテレビ」を開発した。

当初は台風関連データを対象としていたが、途中で東日本大震災が発生し、大震災関連データを対象に加えて事態の推移にリアルタイムで対応する取り組みに方針変更し、データを活用したサービスも開発している。入力と出力をつなぐ部分、具体的には情報源と利用者をつなぐ「情報の交換機」となる部分は、課題未達成であるが、メディアアートとして文化庁メディア芸術祭で受賞という成果もあげている。

「テレビ」は最大にして最強のメディアとのことで、テレビを最終目標とした計画を維持しているが、今後はテレビというメタファーに囚われず、自由な発想で研究を進めてもらいたい。

#### 6. 主な研究成果リスト

##### (1) 論文(原著論文)発表

1. Asanobu KITAMOTO, Takeshi SAGARA, "Toponym-based Geotagging for Observing Precipitation from Social and Scientific Data Streams", Proceedings of the 2012 ACM Workshop on Geotagging and Its Applications in Multimedia, GeoMM'12 (co-located with ACM Multimedia 2012), Liangliang Cao, Gerald Friedland (編), pp. 23-26, ACM, 2012年11月 (in

English)
2. Asanobu KITAMOTO, "Lessons Learned from Data Management Activities after the Great East Japan earthquake in March 2011", Data Science Journal, (in press), 2013.
3. 北本 朝展, "大規模マルチメディアデータの統合と検索による気象イベントのモニタリング", 映像情報メディア学会誌, Vol. 66, No. 11, pp. 907-912, 2012 年 11 月
4. 北本 朝展, "デジタル台風:気象現象のリアルタイム認識・理解に向けた大規模データの統合と分析", 電子情報通信学会技術報告, Vol. 111, No. 222, pp. 5-5, 2011 年 10 月 (招待講演)
5. 北本 朝展, "センサデータとソーシャルメディアの統合によるリアルタイム状況認識の可能性", 電子情報通信学会ヒューマンプロブ研究会公開シンポジウム「ヒューマンプロブの新たな展開」, 2012 年 11 月 (招待講演)

(2)特許出願

研究期間累積件数:0 件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- GeoNLP: <http://agora.ex.nii.ac.jp/GeoNLP/>
- デジタル台風: <http://agora.ex.nii.ac.jp/digital-typhoon/>
- 東日本大震災アーカイブ: <http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/>
- 311 メモリーズ: <http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/311memories/>
- ふってきったー: <http://agora.ex.nii.ac.jp/futtekitter/>
- デジタル台風 TV: <http://digital-typhoon.tv/>
- 第 16 回文化庁メディア芸術祭アート部門審査委員推薦作品「311 メモリーズ」

# 研究報告書

## 「圧縮データ索引に基づく巨大文書集合からの関連性マイニング」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 坂本 比呂志

### 1. 研究のねらい

データ圧縮の研究は、1948 年のクロード・シャノンの論文から始まったといわれている。この論文によって情報理論が創始され、文字の生成確率によって定まるエントロピーと呼ばれる概念によって、データ圧縮を定量的に評価できるようになった。その後、ユニバーサル符号などのアルゴリズムの登場で、データ圧縮の研究が盛んになった。本研究では、大量データ時代の新しいデータ圧縮を実現するための基盤技術の創出を目指す。通常データ圧縮といえばデータサイズを小さくし、ネットワークのスループットや記憶領域の消費量を改善することが目的であると考えられているが、1990 年代の初め頃にはじまった先駆的な研究により、データ圧縮の分野に新しい潮流が生まれている。すなわち、データを圧縮することで様々な計算を高速化する「圧縮情報処理」である。高速化可能な計算としては、圧縮された状態の文書群に対する全文検索や部分文書の高速復元などがあり、これらの技術がさまざまな分野で応用されている。しかし、これらの圧縮情報処理は、静的で比較的小規模のデータに対しては適用可能であるが、ストリームデータのように大量で変化し続けるデータへの適用は困難であった。また現代は、圧縮が必要とされている大規模データ群、高速処理のためのデータ構造とアルゴリズム、圧縮情報処理を可能にする高性能なハードウェアのすべてが揃っている時代でもある。そこで本研究では、圧縮情報処理を大規模データストリームに適用可能な新しい技術を実現し、大規模データゆえに発展が阻害されている分野へこの技術を輸出することで、圧縮情報処理の社会応用を目指す。特に、大規模な計算機環境を利用することが困難な分野や研究課題に対する貢献を目標とし、「データは圧縮して利用するものである」という理念を世界に浸透させる。

### 2. 研究成果

#### (1) 概要

これまでの圧縮情報処理は、BWT や接尾辞配列と呼ばれる文字列のソートに基づく手法が主流で、これらの手法では、同じ接尾辞を持つ部分文字列同士が近くなるように原データを並べ替えるので、圧縮率はよい反面、前処理のためのコストが大きく、また、データが動的に変化するような場合には適用できない。一方、文字列圧縮の世界では、2000 年に提案された文法圧縮と呼ばれる手法が注目を集めている。これは、データを一意に導出する文脈自由文法(CFG)の構文木を構築し、共通の部分木を削除することで圧縮を達成する手法である。文法圧縮はその名が示すとおり、自然言語の構文木を模倣したもので、本来文法構造を持たない DNA シークエンスや文字の羅列に対しても擬似的な文法を構築することで、任意の文字列の圧縮を可能とする。与えられた文字列に対して最小の CFG を構築する問題は、NP-困難と

いう現実的な時間では計算できないと考えられている問題のクラスに含まれるが、ほぼ最適解を線形時間で計算できるアルゴリズムが 2002～2003 年にかけて本さきがけ研究者を含む複数の研究者によって発見された。文法圧縮は、それまでの主要な圧縮法と比較して圧縮率は劣るものの、データの追加に頑健であり、省スペースな領域でも動作するという特徴を持っている。したがって、それまではメモリ上で直接扱うことができなかった規模のデータを圧縮することが可能となった。本研究では、この文法圧縮をさらに発展させ、全文検索や部分文書展開の機構を組み込んだコンパクトな索引構造を構築し、実証実験によってその性能を確認した[原著論文 2,5]。この索引構造はそれまでにないタイプのものであり、特に、Web のアーカイブのように冗長なデータをノート PC のような非常に制限された環境下で極限まで圧縮することができる。本研究では、このデータ構造とアルゴリズムをさらに拡張し、ストリーム型データに適用可能とした[原著論文 3]。このアルゴリズムにより、メモリ効率を保ったまま、データをオンライン圧縮できるため、ソーシャルネットワークを流れる巨大なデータに対してもダイナミックな検索が可能となった[原著論文 1,4]。このアルゴリズムの応用として、巨大な twitter データに対して、ノート PC 上で索引構築から頻出パターン発見までが可能となることを示した。以上のように、本研究では、文法圧縮の可能性を広げ、大規模データストリームマイニングに対する基盤技術を開発した。

## (2) 詳細

### 研究テーマA 「データ圧縮の理論に基づくコンパクトな索引構造の構築」

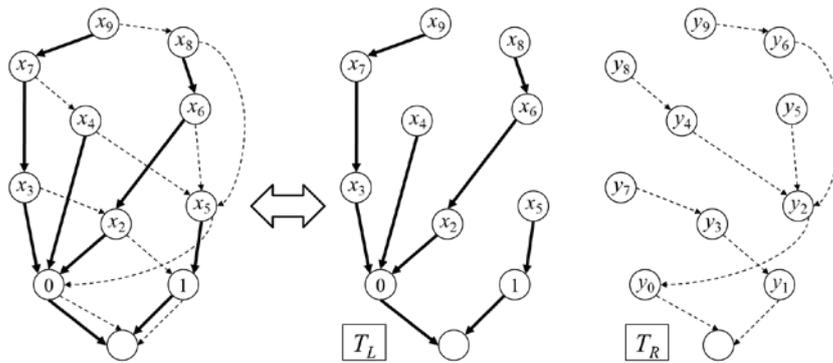
文法圧縮の理論を発展させ、省スペースでストリーム処理可能なオンラインアルゴリズムを構築した。提案手法は、これまでの符号化方と比較して、符号化に要する領域計算量を半分程度に抑えつつ、出力サイズを十分に小さくすることを可能とした。以下の表は、従来手法との出力サイズ(dictionary size)、メモリ使用量(overhead)、計算時間(sec)の比較である。

wikipedia (5.5GB)				
method	dictionary size (MB)	overhead (MB)	time (sec)	
proposed	3,367	5,748	4,903	
dag_vector	10,014	12,511	16,960	
STL vector	9,401	11,898	3,125	
genome (3.2GB)				
method	dictionary size (MB)	overhead (MB)	time (sec)	
proposed	1,819	3,114	2,417	
dag_vector	5,104	6,461	6,359	
STL vector	5,109	6,467	1,576	

### 研究テーマB 「グラフ構造の高速参照と並列化による大規模化」

圧縮データ上で全文検索を実現するためには、補助データ構造を用いた二分探索を実行する必要がある。この補助データ構造が大きすぎる場合、原データサイズを超えてしまい、圧縮索引の意味がなくなってしまう。本研究では、この補助データ構造を含めても文法圧縮のサイ

ズとほぼ変わらないデータ構造を実現した。それが図の文法圧縮の全域木分解である。一般に、文法圧縮はある種の DAG と呼ばれるグラフ構造と等価であり、それは 2 つの全域木に分解できる。木構造は簡潔データ構造と呼ばれる情報理論的下限まで小さく符号化することが可能なため、索引構造全体をコンパクトに表現できる。この符号化手法は本研究で初めて提案された。

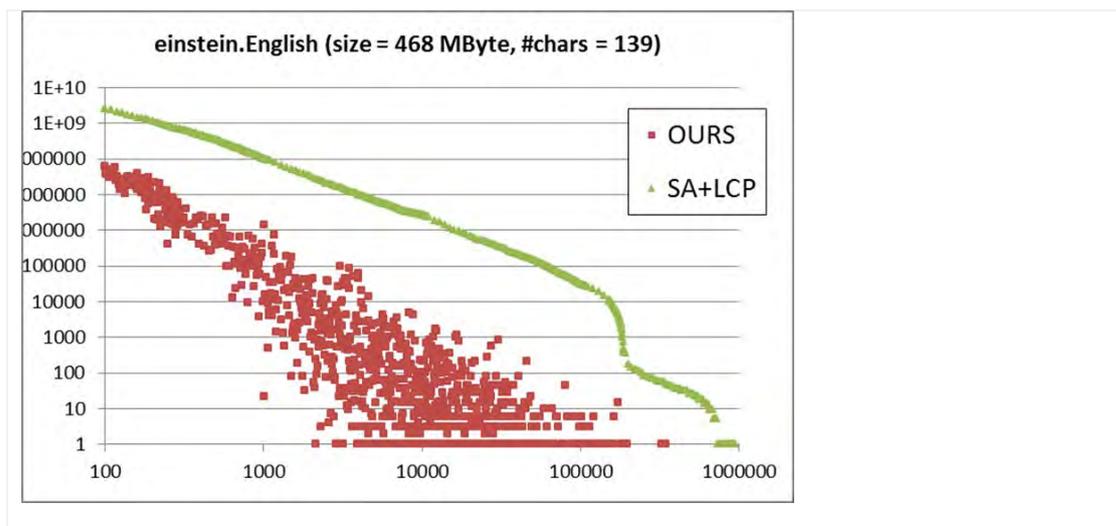


### 文法圧縮の全域木分解

#### 研究テーマC 「圧縮データ索引による重要情報のマイニング」

文法圧縮によって複数のパターンが 1 つの変数(すなわち CFG の非終端記号)として符号化されているとき、その変数を取得して、元のパターンに復元することで、頻出なパターンを近似的に発見しているといえる。本研究では、この文法圧縮上でのパターン発見問題を定式化し、頻出パターンの近似発見問題を理論的に解析した。そしてこのアルゴリズムを大規模実データへ適用することで、実際のパターン発見に適用した。下の表は、Twitter データからの繰り返しパターンの抽出結果である。データサイズは 30GB であり、通常の検索手法ではこの数倍のメモリが必要となるが、本手法では、圧縮後のサイズ 14GB 程度のメモリがあれば実行可能となり、ノート PC 上でこの問題を取り扱うことが可能である。また、グラフの散布図は、パターン検出の近似がどの程度成功しているかを表している。SA+LCP は接尾辞木と高さ配列を用いた厳密解の個数であるが、データサイズの数十倍の領域を必要とする。OURS は本手法で取得したパターンであり、厳密解を約一桁下回るが、線形時間かつ圧縮サイズと同領域で実行可能である。

Text size	Index 構築時間	頻出パターン 発見時間	Index size	出力数 (閾値:80Byte)
約 30 GB	42,956 sec (約12時間)	2,034 sec (約30分)	約 14 GB	55,750,313



### 3. 今後の展開

今後は、本研究で提案した文法圧縮とその応用の利点である、省スペースとオンライン性を最大限に生かした社会応用を目指す。具体的には、SNS に代表されるストリームデータからリアルタイムに情報を取得するための基盤技術として売り出していきたい。また、この分野を発展させるために、「圧縮情報学」のセミナーを開催し、若手の育成に努める。

### 4. 自己評価

本研究の成果は、文法圧縮の理論をほぼ完成させたこと、プログラムを公開して他の研究者のツールを提供したこと、応用例を通して今後の可能性を示せたことである。一方、当初の目標であった、文書群からのマイニングと並列化による大規模化には手が届かなかった。これらは自然言語処理や Hadoop などの専門技術が必要なため、引き続き共同研究として追求していく。

### 5. 研究総括の見解

文字列共起を利用する超大規模なテキストなどの生文書の類似性計算、文書分類、検索に関する新しい技術の開発がテーマである。

さきがけ研究を通じ、文法圧縮の理論をほぼ完成させ、プログラムを公開して他の研究者に提供し、応用例を通して今後の可能性を示しており、評価できる。文書群からのマイニングと並列化による大規模化は果たしていないが、データを圧縮したままでの情報処理の道を切り開いており、評価できる。

「圧縮情報学」を提唱し、この分野を発展させたいとしており、今後を期待している。

### 6. 主な研究成果リスト

#### (1)論文(原著論文)発表

1. Scalable Detection of Frequent Substrings by Grammar-Based Compression,  
M.Nakahara, S.Maruyama, T.Kuboyama, H.Sakamoto,

IEICE Trans. on Information and Systems, in press
2. ESP-Index: A Compressed Index Based on Edit-Sensitive Parsing, S.Maruyama, M.Nakahara, N.Kishiue, H.Sakamoto, Journal of Discrete Algorithms 18:100-112 (2013)
3. An Online Algorithm for Lightweight Grammar-Based Compression, S. Maruyama, H. Sakamoto, M. Takeda, Algorithms 5(2):214-235 (2012)
4. Extracting research communities from bibliographic data, Y.Nakamura, T.Horiike, T.Kuboyama, H.Sakamoto, KES Journal 16(1): 25-34 (2012)
5. Context-sensitive grammar transform: compression and pattern matching, S.Maruyama, Y.Tanaka, H.Sakamoto, M.Takeda, IEICE Trans. on Information and Systems, E93-D(2):219-226(2010)

(2)特許出願

研究期間累積件数:0 件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. Pattern Extraction from Graphs and Beyond,  
H. Sakamoto, T. Kuboyama,  
Book chapter in Multimedia Services in Intelligent Environments,  
Springer, to appear
2. Variable-Length Codes for Space-Efficient Grammar-Based Compression,  
Y. Takabatake, Y. Tabei, H. Sakamoto,  
19th International Symposium on String Processing and Information Retrieval,  
398-410(2012)
3. Scalable Detection of Frequent Substrings by Grammar-Based Compression,  
M.Nakahara, S.Maruyama, T.Kuboyama, H.Sakamoto,  
The 14th International Conference on Discovery Science, 236-246(2011)
4. ESP-Index: A Compressed Index Based on Edit-Sensitive Parsing,  
S.Maruyama, M.Nakahara, N.Kishiue, H.Sakamoto,  
18th International Symposium on String Processing and Information Retrieval,  
398-409(2011)
5. 新聞報道:平成21年12月8日付 毎日新聞「理系白書'09」



# 研究報告書

## 「実社会情報ネットワークからのプライバシー保護データマイニング」

研究タイプ: 通常型

研究期間: 平成21年10月～平成25年3月

研究者: 佐久間 淳

### 1. 研究のねらい

ネットワークベースのサービスや多様な携帯デバイス等の高度化により、多種多様かつ詳細な個人・組織の実社会情報が蓄積されつつある。物品やサービスの購買・販売履歴、ヒトや車両等の精細な地理的移動・行動履歴などの粒度の細かい実社会情報からの知識発見は新しい価値を持つサービスを生み出す源泉となりうるが、知識発見の対象となる個人情報の詳細度が増せば増すほど、それを開示・活用することへの心理的抵抗感は大きくなる。多様な知識資源が計算機にとって活用可能な形式で蓄積されつつある中、プライバシー保護と知識獲得を両立させる技術は知識獲得技術の発展のためには不可欠であるといえる。

このような社会的背景を鑑み、分散秘密情報源からの安全な知識獲得を目指すプライバシー保護データマイニング(privacy-preserving data mining, PPDM)が近年活発に研究されている。PPDMとは、二人(以上)のエージェントがそれぞれ秘密の入力データセットを持つときに、それらが互いにデータ見せ合うことなくデータマイニングを実行し、得られた知識のみの共有を目指す技術である。従来、PPDMでは、入力データはデータベーステーブルに格納されていることを想定し、多くは2エージェント間において、属性単位やレコード単位で秘密情報を定義することが想定されてきた(プライバシーモデル)。しかしながら、「何処へ行き何をしたか」、「何に興味を持ち何を購入したか」、「誰は誰と協力関係にあるか」といった実社会情報から、プライバシー保護を考慮しつつ知識発見を実現するためには、情報の構造、流通経路、情報の開示可能な範囲、計算主体の物理的所在や通信・計算能力等を考慮してプライバシーモデルの定義をし、また知識発見アルゴリズムもそれに応じて設計する必要がある。

本研究では、情報の構造、流通経路、計算主体の性質等を考慮したネットワーク構造によるプライバシーモデルである private graph を中心概念とし、これに基づく実社会情報ネットワークからの安全な知識獲得を実現する多様なアルゴリズム群と、その実行・開発フレームワークである Fairy Ring を構築した。これにより、実社会のネットワーク構造情報が持つ秘匿性を侵害することなく、安全に知識発見および発見知識の流通を実現するための理論的基盤との構築と開発技術の提供を実現した。

### 2. 研究成果

#### (1) 概要

本研究課題の研究テーマは大きく下記の3テーマで計画された。

研究テーマ A プライベートグラフにおける秘密計算

研究テーマ B プライベートグラフにおけるデータ集約・匿名化と開示の理論

研究テーマ C プライベートグラフにおける分散データマイニング開発環境

研究テーマ A については、ノードランキング、ノードラベル予測、オンライン予測等のアルゴリズムを実現し、研究テーマ C において、これらをアンドロイド上で実現可能なシステム Fairy Ring を開発した。研究テーマ B については、ネットワーク構造情報を表現する行列を含む、半正定値行列を、安全に開示(差分プライバシー)するための理論的枠組みと、それを実現するアルゴリズムを提案した。以下、各テーマについて詳細を説明する。

## (2) 詳細

研究テーマ A1. プライベートグラフにおけるランキング問題:

”誰が誰から買った”, ”誰が誰に電話(メール)をした”, ”誰が誰を評価した”, などのリンクや社会的紐帯は、機密性やプライバシーの問題のため、公にされることが好まれない場合が多い。この研究では、秘密のリンクを含むネットワークを対象として、その秘匿性を損なうことなく安全にリンク解析を適用可能な方法を開発し、これまで秘密性のために解析対処とならなかった現実の多様なネットワークからの情報抽出を可能にした。具体的には、この研究で提案したアルゴリズム PrivateRank は Web リンク解析において著名な PageRank 法をプライベートグラフ上で安全に実現する方法を構築した。(受賞1、論文投稿予定)

研究テーマ A1 プライベートグラフにおけるラベル予測問題

このテーマでは、ノードがラベル情報を持つプライベートグラフにおいて、秘密を保護しつつノードラベルを予測するアルゴリズムを提案した。例として、ノードが個人、リンクが物理的な接触関係、ノードのラベルが感染状態であるような感染症ネットワークが考えられる。ラベル予測のための従来の半教師付き学習方法はリンクやラベルの公開を前提としている。このアルゴリズムでは、加法準同型性公開鍵暗号を用いたマルチパーティーコンピューテーションによって半教師付き学習の一方式であるラベル伝播法をプライベートグラフ上で安全に実行する方法を開発し、ラベルとリンクが各ノードの秘密情報であっても、その秘密を保護しつつ、ラベル予測を可能にした。(論文2, 受賞1)

研究テーマ A2 エキスパート間の秘密のアドバイスの基づくオンライン予測

このテーマでは、学習者が複数のエキスパートの忠告(予測)を各時刻において得て、これに基づいて次の時刻の系列の予測を行うオンライン予測問題において、エキスパートの忠告に秘密情報が含まれるために、他のエキスパートや学習者に開示できない状況におけるオンライン予測を実現する方法を構築した。エキスパートが保持する予測情報を他者と共有することができない場合の学習者の予測精度は、それが共有できた場合と比べ悪化するよう思えるが、構築した secure exponential weighting 法では、暗号理論的ツールを利用することによって、予測/損失情報を学習者や他のエキスパートと共有しなくても、それらがすべて共有された場合と同じ予測精度が達成可能であることを理論的に示すことに成功した。このことから、オンライン予測において、エキスパートの忠告のプライバシー保護は、オンライン予測の妨げにならないことを理論的に示した。(論文5、受賞3)

研究テーマ A3 プライベートグラフにおけるラベル予測問題

この研究ではラベル付きプライベートグラフにおいて、ノードラベルが秘密情報である場合に、ラベルやリンクの秘密を保護したままラベルを予測するプロトコルを構築した。例として、ノードが個人、リンクが接触、ノードのラベルが感染状態である感染症ネットワークにおける感染状態予測精度が挙げられる。ラベル予測のための従来の半教師付き学習方法はリンクやラベルの公開を前提としていたが、本方式では、加法準同型性公開鍵暗号を用いた秘密計算を半教師付き学習の一方式であるラベル伝播法に導入して、プライベートグラフ上で、安全にラベル予測が実行可能になった。(論文1、受賞2)

#### 研究テーマ A4 プライベートグラフにおける安定結婚問題

安定結婚問題とは、男女同数のグループについて、それぞれの選好に基づいた安定した一対一マッチングを求める問題である。ここで、選好とは、ある人の好みに基づく異性のグループに対する順位付けであり、安定したマッチングとは、どの二つのマッチングを組み換えても、組み換えられた男女の選好の順位が低くなることを指す。この選好を表す情報は、個々人は個々人のセンシティブな情報の一つであるケースも多くプライバシーの問題が生じる。この研究では、全員の選好のプライバシーを保護しつつ、安定マッチングのみを計算可能なアルゴリズムを構築した。(発表1、論文投稿予定)

研究テーマ A における代表的な成果は以上4つであるが、これに加え、この領域において以下のアルゴリズム開発を行った。

A5 PPDM におけるパーティ同士の結託を防ぐ基礎的プロトコル (論文4)

A6 複数プライベートグラフにおけるノードランキング問題 (論文2)

#### 研究テーマ B プライベートグラフにおけるデータ集約・匿名化と開示の理論

##### B1 半正定値行列の差分プライバシーと低ランク近似

ソーシャルネットワークや、ユーザ・アイテムの購入関係を表すネットワーク、人の移動履歴が構成するネットワークなど、サービスを行う特定の事業者が収集可能なネットワークについて、これが外部公開可能であれば有益なデータ解析が可能であると考えられるが、その公開がプライバシー侵害になる可能性がある。

研究テーマ A はプライベートグラフ上のノードが計算主体となり、各ノードが分散し保持する秘密情報を入力として、秘密計算を行うアルゴリズム群を中心とした研究を行った。これに対して、研究テーマ B は、このようなプライベートグラフに関する情報を収集するエンティティがプライベートグラフ自体を外部に公開する際に発生するプライバシーの問題に関する理論的考察と、安全に公開するためのアルゴリズムの開発を行った。

具体的には、プライベートグラフを半正定値行列として表現し、これを公開するとき、行列自体ではなく、その低ランク近似生成し、これに、そのランクに比例するスケールでノイズを加えることによって、差分プライバシーとよばれる安全性を保証できることを理論的に示した。またそのようなランダム化を実現するためのアルゴリズムを構築し、プライベートグラフを安全に公開するための方法を示した。(発表3、4、論文投稿中)

## B2 秘密計算における差分プライバシーのための指数メカニズム

研究テーマ A で開発したような様々なアルゴリズムの秘密計算は、入力を秘匿しつつある定められた計算を行い、その出力のみを共有することを可能にするが、出力は入力情報が正しく反映されるために、秘密計算の出力から入力情報が漏れる場合が存在する。この研究では、秘密計算の出力に差分プライバシーを保証するメカニズムを構築した。具体的には、差分プライバシーを与えるメカニズムのクラスを一般的に捉えることができる指数メカニズムの概念を用いる。具体的には、ある秘密計算を指数メカニズムになるよう変形し、その指数メカニズムとしての振る舞いから、与えられる差分プライバシーを評価する。本稿では、指数メカニズムの実行に必要なサンプリングを加法準同型暗号を用いた重み秘匿紛失ルーレットにより行う方法について述べる。(発表2、論文投稿予定)

## 研究テーマ C プライベートグラフにおける分散データマイニング開発環境 Fairy Ring

研究テーマ C では、研究テーマ A で開発したアルゴリズム群のための開発環境および、これを実際にサービスとして 提供するための実行環境として、スマートフォンを利用した秘密計算開発・実行環境 Fairy Ring を構築した。このフレームワークの特徴としては、従来の秘密計算の応用シナリオでは、計算能力が高く、常時接続・常時稼動を前提とした、複数の対等な高性能サーバ間での実行を想定して設計されてきたが、プライベートグラフ上での計算にあるように、日常生活におけるカジュアルな用途での秘密計算はほとんど検討されてこなかった。近年のスマートフォンは、計算性能や通信速度に優れ、従来の携帯端末では、現実的な計算時間で実行困難であった秘密計算が、実現可能になりつつことから、このフレームワークでは、準同形性公開鍵暗号を利用した少数の基礎的な秘密計算をビルディングブロックとし、これを組み合わせることで、研究テーマ A で開発したアルゴリズム群を含む、様々な個人間の意思決定を簡単に可能にする秘密計算を実装した。(発表1、5、論文投稿予定)

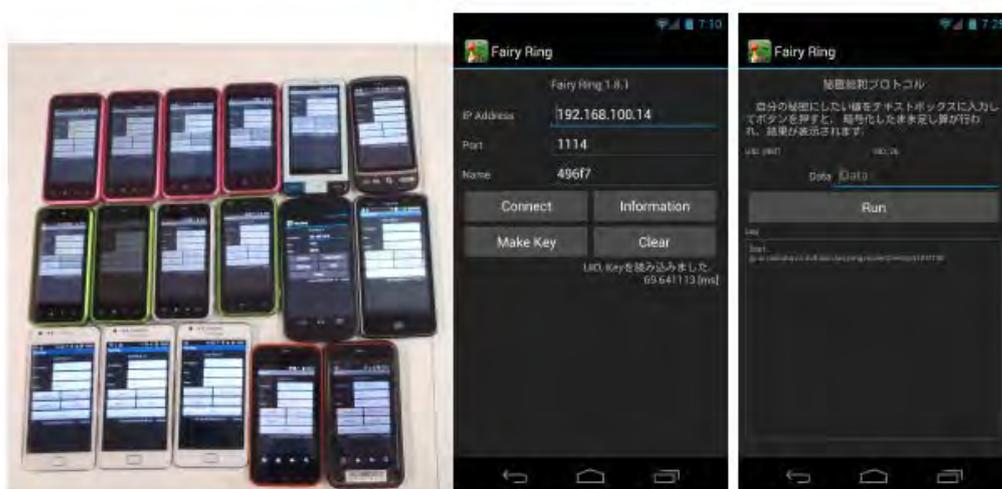


Figure 1 Fairy Ring の動作画面

### 3. 今後の展開

プライベートグラフを中心概念として、これにまつわる多様なアルゴリズム群とその Android 上での実行・開発環境 Fairy Ring を構築した(研究テーマ A および研究テーマ C)。Fairy Ring については、これをサービスとして利用するユーザの観点について、

- Android Market での公開
- ユーザーインターフェースの改良
- 利用可能な秘密計算の種類を増強

Fairy Ring を開発環境として利用する開発者の観点について、

- フレームワークの整理
- ライブラリの増強

を今後行う予定である。

広くユーザに秘密計算に触れる機会を提供することで、新たなニーズを探りつつ、秘密計算が提供できる新しい知識やサービスや価値について考察を深め、今後の研究開発に役立てたい。

研究テーマ B では、プライベートグラフの安全な開示を実現するための理論の構築を目指した。理論的には新規性の高い方式を提案した。一方、理論を実現するためのアルゴリズムはマルコフ連鎖モンテカルロ法に頼るため、大規模グラフにおいては計算時間に問題がある点が課題である。この点についての解決をはかるとともに、論文公表を行うことが今後の課題となる。

大局的な視点からは、ストリームデータの計算モデルにおける高速な秘密計算モデルの構築を構想している。比較的少数のエンティティを対象とし、複雑であるがデータ量自体は比較的少ないグラフデータとは異なり、大規模データを秘密計算に適用するための新しい計算モデルを開発したい。

### 4. 自己評価

この研究は、全体としてネットワーク構造情報にまつわるプライバシーを安全に利用するための一連の理論と技術を提供とすることをねらいとしていた。この点について、プライベートグラフという基礎概念を元に、主にネットワークマイニングや多数意思決定のための多くのアルゴリズムをマルチパーティー秘密計算として実行するためのプロトコル群を提供できたことについては、機会学習分野およびセキュリティ・プライバシー分野の両方にとって、貢献が大きかったと評価している。また、現実世界でプロトコルを動作させ、またプロトコル開発を容易にするためのフレームワークである Fairy Ring の開発は、目的からも計算方式からも、秘密計算の開発環境としてユニークであり、今後これを利用した様々な応用を見だし、秘密計算のコモディティ化・サービス化への道を開きたい。

プライベートグラフ自体の安全な公開のための理論的枠組みについては、新しい方法論を用いるテーマだったこともあり、さきがけ期間中の論文発表に至っていないが、広い応用範囲を持つ基礎技術であり、今後は情報な安全な公開のための理論研究を進める上では、研究者自身としては重要な成果として認識している。現在論文投稿中であり、今後、この領域での貢献を目指す。

### 5. 研究総括の見解

プライベートなデータを扱うデータマイニングにおいて、プライバシーを確保するプライベート・グ



ラフ・モデルの提案、およびその手法についての研究である。

ネットワーク環境でプライバシーを安全に利用するという点について、プライベートグラフという基礎概念を元に、主にネットワークマイニングや多人数意思決定のための多くのアルゴリズムをマルチパーティー秘密計算として実行するためのプロトコル群を提供したことは評価できる。

また、現実世界でプロトコルを動作させ、またプロトコル開発を容易にするためのフレームワークである Fairy Ring を開発し、これを利用して様々な応用を見いだす体制を開発しており、今後に期待したい。

## 6. 主な研究成果リスト

### (1) 論文(原著論文)発表

- |  |
|--|
| 1. Hiromi Arai, Jun Sakuma: Privacy Preserving Semi-supervised Learning for Labeled Graphs. Lecture Notes in Computer Science 6911 (Machine Learning and Knowledge Discovery in Databases – European Conference, ECML PKDD 2011), pp. 124–139. |
| 2. 森井正覚, 佐久間淳, 佐藤一誠, 中川裕志: 統合したグラフのプライバシー保護リンク解析. 情報処理学会論文誌, Vol.50 TOD 4(2). pp.52–60. 2011.  |
| 3. Jun Sakuma, Shigenobu Kobayashi, Large-scale k-means clustering with user-centric privacy-preservation, Knowledge and Information Systems, 2010, No. 2, pp. 253–279.  |
| 4. Bin Yang, Hiroshi Nakagawa, Issei Sato, Jun Sakuma: Collusion-resistant privacy-preserving data mining, Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 483–492.                    |
| 5. Jun Sakuma, Hiromi Arai, Online Prediction with Privacy, Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 935–942.   |

### (2) 特許出願

研究期間累積件数: 0件

### (3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

#### 主要な学会発表

(1) 照屋唯紀, 川本淳平, 佐久間淳, 強い安全性を仮定したスマートフォン向けマルチパーティー秘密計算フレームワーク, 2013年暗号と情報セキュリティシンポジウム(SCIS2013), CDRROM 予稿集 pp. 1–8.

(2) 照屋唯紀, 川本淳平, 佐久間淳, 指数メカニズムによる秘密計算の出力における差分プライバシーの保証, 2013年暗号と情報セキュリティシンポジウム(SCIS2013), CDRROM 予稿集 pp. 1–8.

(3) 佐久間 淳, 半正定値行列の差分プライバシー, 第15回情報論的学習理論ワークショップ(IBIS 2012)(第11回IBISML研究会), pp. xx–yy.

(4) 佐久間 淳, スペクトラル差分プライバシーに基づくプライバシー保護推薦アルゴリズム, 第26回人工知能学会全国大会, 2012, pp. CDRROM 予稿集 pp. 1–4.

(5) 青木 良樹, 佐久間 淳, 照屋 唯紀, どこでも秘密計算フレームワーク Fairy Ring マルチメ



ディア、分散、協調とモバイル DICOMO2012 シンポジウム, pp. 1618–1627.

#### 受賞

- (1) 2009 年度(第 23 回)人工知能学会全国大会優秀賞「秘密のリンク構造を持つグラフのリンク解析」
- (2) 情報論的学習理論と機械学習 (IBISML2010) 研究会 Honorable Mention「ラベル付きグラフに対するプライバシーを保護した半教師付き学習法」
- (3) 情報論的学習理論と機械学習 (IBISML) 2010 年度研究会賞 電子情報通信学会 第一種研究会「情報論的学習理論と機械学習研究会」「オンライン予測におけるプライバシー保護」
- (4) コンピュータセキュリティシンポジウム 2012 優秀論文賞, 「Bloom Filter を用いた積集合サイズのベイズ推定とそのプライバシー保護疫学調査への応用」
- (5) 平成 24 年度 日本データベース学会上林奨励賞

#### プレスリリース

- (1) 秘密計算による化合物データベースの検索技術, 2011.11.1

[http://www.aist.go.jp/aist\\_j/press\\_release/pr2011/pr20111101/pr20111101.html](http://www.aist.go.jp/aist_j/press_release/pr2011/pr20111101/pr20111101.html)

# 研究報告書

## 「密度比推定による大規模・高次元データの知的処理技術の創生」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 杉山 将

### 1. 研究のねらい

社会・経済構造の急激な変化, 高齢化社会, 環境問題, 領土問題などへの迅速な対応が叫ばれている昨今, コンピュータで処理すべき情報量は爆発的に増加している. 情報処理の立場からは, この問題は「データの数」, 「データの次元」, 「データ構造の複雑さ」が著しく増加していると解釈できる. このような背景のもと, データ処理に統計的な手法が積極的に用いられるようになってきており, 特に**機械学習**と呼ばれる分野が学术界・産業界から大きな注目を集めている. 本研究のねらいは, 複雑な構造を持つ超高次元かつ莫大な量のデータから, その背後に潜む有用な規則を自動的に見つけ出す機械学習技術の新しいパラダイムを創生することである.

機械学習の目的は, 与えられたデータからその背後に潜む一般的な規則を自動的に獲得することである. 例えば, キーボードやマウスなどを使わずに脳波で直接コンピュータを操作しようというブレインコンピュータインターフェースにおいては, 人間の脳波のパターンと人間の意志(カーソルを右や左に動かす等)との関係が学習の対象となる. このような機械学習の研究において, データの統計的な振る舞いを調べることは学習の効率を向上させるために非常に重要である. しかし一方では, データの確率分布そのものを推定することは避けるべきであるという考えがある. なぜならば, データの確率分布の推定は学習問題そのもの(上記の例では脳波と意思とのマッピングの学習)よりも本質的に難しい問題であり, 確率分布の推定を経由することにより規則の学習の精度が大きく低下してしまうからである. この原理は数学者 Vapnik により提唱され, この原理に基づいたサポートベクターマシンとよばれるパターン認識アルゴリズムが, 非常に優れた性能を示している.

本研究では, この Vapnik の原理を応用し, 重点サンプリング, 確率分布比較, 相互情報量推定, 条件付き確率推定などの様々な機械学習タスクを, 「**確率密度比**」の推定により統一的に解決する枠組みを提案する. そして, 密度比推定に基づいた新しい機械学習アルゴリズム群を開発し, その数理的な性質を明らかにする. 更に, 信号画像処理, ロボット制御, 脳波解析, 自然言語処理, 生命情報解析などの分野に応用する.

## 2. 研究成果

### (1) 概要

Vapnik の原理に従えば、確率密度比の推定は確率分布の推定より容易に行うことができる(図1)。なぜならば、各密度が分かればそれらの比がわかるが、比が分かっても各密度は特定できないからである。そのため、密度比推定を通して機械学習を行えば、密度推定を用いるよりも精度が良くなると期待される。



図1：密度推定と密度比推定

本研究の開始にあたって、重点サンプリング、確率分布比較、相互情報量推定、条件付き確率推定などの様々な機械学習タスクが、確率密度でなく確率密度比のみを推定することにより解決できることを示した。そのもとで、研究テーマ A「確率密度比の直接推定手法の開発」、研究テーマ B「密度比推定の数理的性質の解明」、研究テーマ C「密度比推定に基づく機械学習のアルゴリズム開発」、研究テーマ D「実応用」の研究を行った。

### (2) 詳細

#### 研究テーマ A「確率密度比の直接推定手法の開発」

まずは、カルバック・ライブラー情報量に基づく密度比の直接推定法 KLIEP を開発し、密度比推定という新しい研究分野を切り開いた。その後、二乗損失に基づく簡便な密度比推定法 uLSIF を開発した。この手法により、密度比推定にかかる計算コストを大幅に軽減することができるようになり、様々な機械学習アルゴリズムへの応用可能性が広まったとともに、理論解析を行うための数理的な基礎も築いた。また、uLSIF 法に基づき、高次元空間での密度比推定に次元削減を組み合わせる新しいアルゴリズムも開発した(論文3)。

#### 研究テーマ B「確率密度比推定の数理的性質の解明」

上記の KLIEP 法と uLSIF 法が、パラメトリック推定とノンパラメトリック推定の両方の枠組みにおいて、理論的に最適な近似精度を達成していることを証明した。更に uLSIF 法は数値的安定性の意味でも最適な手法になっていることを示した。また、KLIEP 法と uLSIF 法を含む、これまでに提案されてきた全ての密度比推定法を含む統一的な枠組みを考案し、各手法の相互関係を明らかにした(論文2)。

#### 研究テーマ C「確率密度比に基づく機械学習のアルゴリズム開発」

大きく分けて、重点サンプリング、確率分布比較、相互情報量推定、条件付き確率推定に関する4つの研究を行った。「重点サンプリング」では、データを生成する確率分布が変化する非定常環境下での適応学習アルゴリズム(図2)とモデル選択手法、能動学習アルゴリズムなどを開発した。「確率分布比較」では、異常検出(図3)、二標本検定(論文4)、時系列の変化点検知などのアルゴリズムを開発した。「相互情報量推定」では、独立性検定、特徴選択、特徴抽出、独立成分分析、因果推定、オブジェクト適合、クラスタリングなどのアルゴリズムを開発した(論文1, 図4)。「条件付き確率推定」では、条件付き密度推定、確率的パターン認識(論文5)、マルチタスク学習、マルチラベル学習などのアルゴリズムを開発した(図5)。

## 研究テーマ D「実応用」

重点サンプリングに基づく機械学習手法を、音声からの話者認識、日本語文書の単語分割、ブレインコンピュータインターフェース(図6)、ロボット制御などに応用した。確率分布比較に基づく機械学習手法を、製鉄プロセスの異常診断、光学部品の品質検査、画像中の注目領域の抽出、動画やツイッターからのイベント検出などに応用した。相互情報量推定に基づく機械学習手法を、遺伝子解析、製鉄プロセスの解析、医療画像処理、モーションキャプチャデータの解析などに応用した。条件付き確率推定に基づく機械学習手法を、製鉄プロセスデータの可視化、顔画像からの年齢認識、加速度センサからの行動認識などに応用した。顔画像からの年齢認識装置は2012年春に NEC ソフトより商品化された。

## その他「広報活動」

本研究で得られた成果をまとめた英語専門書を MIT Press 社より出版した(著書1, 図7)。応用分野を開拓するため、国内外の企業・研究所・大学にて60回以上の招待講演を行った。

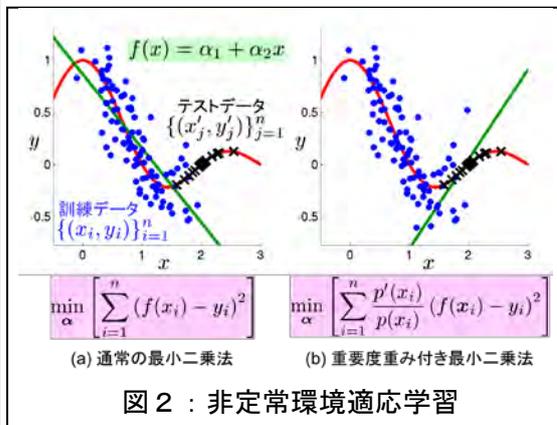


図2：非定常環境適応学習

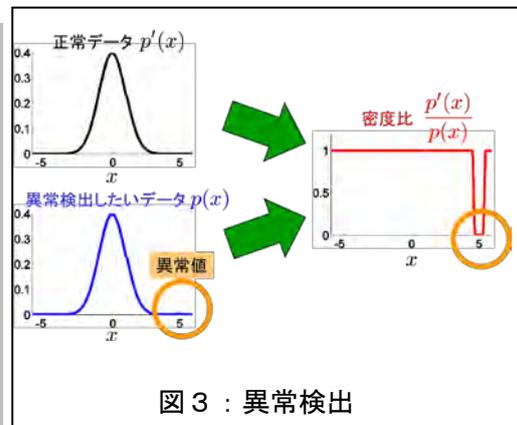


図3：異常検出

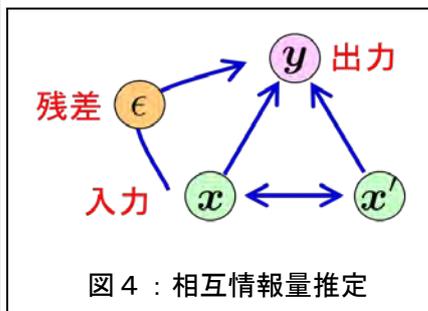


図4：相互情報量推定

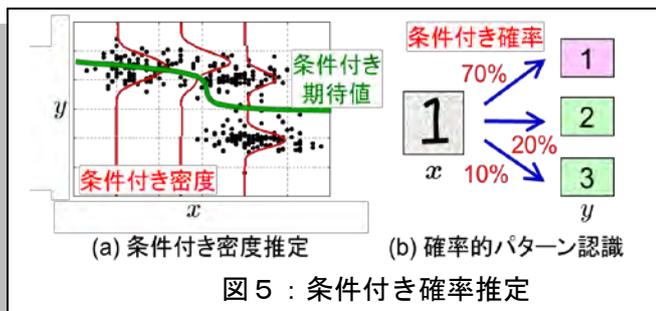


図5：条件付き確率推定

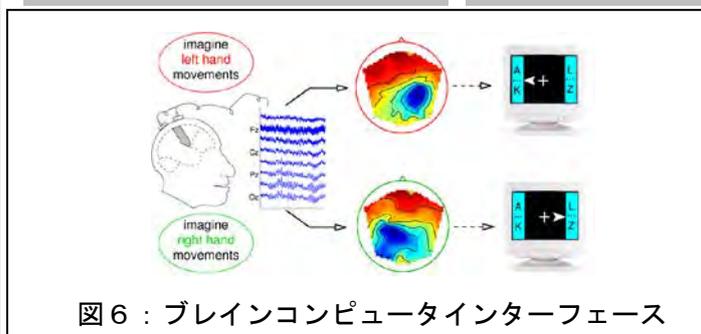


図6：ブレインコンピュータインターフェース

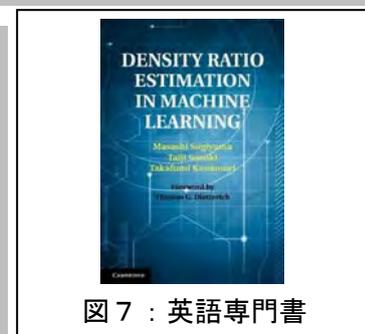


図7：英語専門書

### 3. 今後の展開

機械学習技術はこの数年で一気にコモディティ化が進み、研究者だけでなく企業の技術者にもそのキーワードが認知されるようになった。そのため、さきがけ研究期間中に多くの企業と共同研究をする機会に恵まれた。ビッグデータ時代のこれからも、機械学習技術の重要度は益々高まっていくものと考えられるため、今後も機械学習技術の実用化を更に加速させていきたい。

情報技術の応用パートナーといえ、数年前まではIT系の企業が多かったが、近年は材料メーカーや物流など多岐にわたる企業が機械学習技術に興味を持ちつつある。そこで、特に非IT系の企業とのコラボレーションを加速させて、新しい応用分野を開拓していきたい。その際には、基礎的な技術の普及が重要なカギを握るため、機械学習の入門書を執筆したり、チュートリアルをしたりという広報・教育活動も積極的に行いたい。

このように色々な分野の研究者や技術者が入り混じる機械学習分野における研究は、もはや個人で行うのは不可能だと感じられるほど裾野が広がってきている。そこで、基礎から応用まで幅広くカバーできるチームを作って研究をすすめていきたい。その際、現行技術の改良だけでなく、次世代の独創的なパラダイムを生み出せるよう地道な基礎研究をしっかりと行なっていきたい。

### 4. 自己評価

本研究プロジェクトでは、密度比推定という新技術を軸に、その技術が応用できる実問題を探すというスタイルで研究を進めてきた。幸い多くの企業に興味を持ってもらえ、多数の応用事例が見つかった。しかし、技術に基づくボトムアップ型の研究スタイルであるため、基本的には解ける問題を探して解いていたということになる。ビッグデータ時代には、これまでの技術では到底解決できないような難問が次々と現れることが予想されるため、世の中の風潮としては、応用に基づくトップダウン型の研究スタイルが好まれるようになってきていると感じている。

応用に基づいて研究が発展していく潮流そのものを否定するものではないが、このままでは技術に基づく研究がないがしろにされてしまうのではないかという懸念を強く感じている。情報技術は、色々な応用分野で活用されることによって、その研究成果を実世界に還元できるものであるが、情報技術そのものの基礎研究を軽視すべきではないと強く感じる。

本研究の成果は、密度比を使って簡単に解ける問題を見つけて解いたと解釈できる。しかし、問題がもともと簡単であったわけではなく、密度比を用いることによって、難しかった問題を簡単な形に変換できたため、結果としてうまく解けたのである。このような研究成果が得られたのは、密度比推定という技術が軸になっている研究計画に対して、さきがけの強いサポートを受けられたからである。応用研究がより重要視される中、今後も一定量の情報学の基礎研究をサポートする体制が存続して欲しいと心より願う。

一方、自分の作った理論を誰かが見出して応用してくれると信じている理論研究者が、自分の世界に閉じこもってしまっているのも事実である。従って、理論と応用の橋渡しをより積極的に進めていく必要がある。そのためには、前述したように基礎から応用までをカバーする研究チームを作ることが重要であると考えている。この理想的な研究体制が実現できるよう、更に自己研鑽を続けていきたい。

### 5. 研究総括の見解



データ解析の分野で、密度比推定という枠組みをさまざまな応用分野に適用可能にするという研究である。

この研究では、密度比推定という新技術を軸に、その技術が応用できる実問題を探すというスタイルで研究を進め、多くの企業と共同で多数の応用事例を解決している。また、この技術を普及させるための教科書も出版している。着実に成果をあげていることを評価したい。

## 6. 主な研究成果リスト

### (1)論文(原著論文)発表

1. Sugiyama, M. Machine learning with squared-loss mutual information. Entropy, vol.15, no.1, pp.80-112, 2013.
2. Sugiyama, M., Suzuki, T., & Kanamori, T. Density ratio matching under the Bregman divergence: A unified framework of density ratio estimation. Annals of the Institute of Statistical Mathematics, vol.64, no.5, pp.1009-1044, 2012.
3. Sugiyama, M., Yamada, M., von Büna, P., Suzuki, T., Kanamori, T., & Kawanabe, M. Direct density-ratio estimation with dimensionality reduction via least-squares hetero-distributional subspace search. Neural Networks, vol.24, no.2, pp.183-198, 2011.
4. Sugiyama, M., Suzuki, T., Itoh, Y., Kanamori, T., & Kimura, M. Least-squares two-sample test. Neural Networks, vol.24, no.7, pp.735-751, 2011.
5. Sugiyama, M. Superfast-trainable multi-class probabilistic classifier by least-squares posterior fitting. IEICE Transactions on Information and Systems, vol.E93-D, no.10, pp.2690-2701, 2010.

### (2)特許出願

1.

発明者：植木 一也, 伊原 康行, 杉山 将.

発明の名称：目的変数算出装置, 目的変数算出方法, プログラムおよび記録媒体

出願人：NEC ソフト, 東京工業大学

出願日：2009年9月28日

出願番号：特願 2009-221989, 特開 2011-070471

2.

発明者：平田 文英, 河原 吉伸, 杉山 将.

発明の名称：パターン自動抽出方法およびパターン自動抽出システム

出願人：JFEスチール株式会社

出願日：2010年5月25日

出願番号：特願 2010-119743, 特開 2011-247696

### (3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

著書1: Sugiyama, M., Suzuki, T., & Kanamori, T. Density Ratio Estimation in Machine Learning, 344 pages, Cambridge University Press, Cambridge, UK, 2012

受賞1: 2010年5月13日 電子情報通信学会 PRMU 研究会 研究会奨励賞



受賞2:2010年5月14日 人工知能学会 DMSM 研究会 研究会優秀賞

受賞3:2011年6月2日 情報処理学会 長尾真記念特別賞

受賞4:2011年12月16日 日本神経回路学会 論文賞

受賞5:2012年4月14日 船井情報科学振興財団 船井学術賞

受賞6:2012年6月19日 電子情報通信学会 IBISML 研究会 研究会賞ファイナリスト

# 研究報告書

## 「インフルエンザ感染伝播のデータ同化モデルによる解析・予測技術」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 鈴木 秀幸

### 1. 研究のねらい

新型インフルエンザのパンデミック発生回避や被害軽減のための方策を検討する際には、感染伝播モデルによる解析・予測が有効であると期待されるが、単なる数値シミュレーションではモデルと現実との乖離が問題となる。本研究では、データ同化技術を導入することにより、現実のデータとの整合性の取れたシミュレーションを実現し、感染伝播モデルによる解析・予測を行うための数理的基盤技術を開発することをねらいとした。

### 2. 研究成果

#### (1) 概要

本研究では、インフルエンザ感染伝播のデータ同化シミュレーションに関する研究と、インフルエンザ感染伝播の多様性に関する数理モデル研究を行った。データ同化シミュレーションに関する研究においては、パーソントリップ調査に基づく東京都市圏感染伝播シミュレータから得られる新規感染者数をデータとして用い、(実効的な)基本再生産数の推定を行った。この結果により、データ同化シミュレーションが、対応策等の効果の推定に利用できる可能性を示した。また、インフルエンザ感染伝播の多様性に関する数理モデル研究においては、個体の多様性、感染症の多様性、移動・接触の多様性を考慮した数理モデルを構築し、これらの数理モデル上で、感染ダイナミクスの解析やワクチン接種戦略等に関する検討を行った。以上のように、データ同化技術を導入するとともに、インフルエンザ感染伝播の多様性を考慮した数理モデル研究を行うことにより、モデルと現実との乖離の問題に対処するための数理的基盤技術の開発を行った。

#### (2) 詳細

##### 研究テーマ A 「東京都市圏感染伝播シミュレータを用いたデータ同化シミュレーション」

本テーマにおいては、東京都市圏感染伝播シミュレータにおける数値計算結果から得られた感染者数の推移のデータを用いて、SEIR (Susceptible-Exposed-Infected-Recovered) モデルによるデータ同化シミュレーションを行った。

東京都市圏感染伝播シミュレータ(論文 1)は、Halloran et al. (Vaccine 20, 3254, 2002)等で用いられた状態遷移モデルに基づくマルチエージェントシステムであり、各個体に対してそれぞれの感染状態の遷移を考える。各個体はパーソントリップ調査データの各個人の動きに従って異動させて、時々刻々の個体間の接触を考える。本研究で用いた移動データは、東京都市圏交通協議会による平成 10 年首都圏パーソントリップ調査に基づくものであり、東京大学空間情報科学研究センターの人の流れプロジェクトにおいて時空間データに変換されたものである。

データ同化シミュレーションは、Saito et al. (Int. Conf. Fusion, 2011)に従い、SEIR モデルを用いて(実効的な)基本再生産数の推定を行った。その結果、データに追従した SEIR モデルのシミュレーションが行われることが確認されたほか、基本再生産数が 1 を切るときに新規感染者数がピークとなることが確認された。さらに、東京都市圏感染伝播シミュレータにおいて外出規制をかけた場合についてもデータ同化シミュレーションを行った。規制日数が不十分であると規制解除後に感染が再拡大するが、規制が不十分であることが基本再生産数の推定値に現れていることを確認した。この結果により、対応策の効果が十分であるか判定に利用できる可能性を示した。

#### 研究テーマ B 「複数宿主 SIR モデルにおける動的ワクチン接種戦略」

本テーマにおいては、個体の多様性を考慮するため、宿主を感染率・致死率等のパラメータの異なる複数のタイプに分けた SIR モデル上において、ワクチン接種の動的な最適接種戦略を検討した。また、目的関数として、感染者数最小化と死亡者数最小化を検討した。本研究では、接種可能なワクチンの全期間にわたる総量が制限されているほか、供給速度にも制限があることを仮定している。未成年、成人、高齢者の3タイプを想定したモデル上で最適化計算を行った結果、目的関数やパラメータの違いにより最適戦略が変化することを確認した。得られた最適解の中には、非自明な動的変化が見られた。

#### 研究テーマ C 「複数宿主適応的ネットワーク上の感染ダイナミクスの理論的解析」

本テーマにおいては、個体の感染状態に応じてネットワーク構造が変化するモデルを検討した(論文2)。これは感染者との接触を避ける行動と対応している。また、前テーマと同様に、複数タイプを考慮して、同タイプ内での感染確率および接触確率を定め、微分方程式モデルを定式化した。さらに、この微分方程式モデルから基本再生産数を導出した。その結果、感染確率と接触確率が互いに同じ方向性を持っているときに感染が拡がりやすく、異なる方向性を持っているときに感染が拡がりにくいことが理論的に明らかになった。

#### 研究テーマ D 「新型・季節型同時流行の SIR モデルにおける動的ワクチン接種戦略」

本テーマにおいては、新型インフルエンザと季節性インフルエンザの同時流行を想定した SIR モデル上において、ワクチン接種の動的な最適接種戦略を検討した。本テーマにおいても目的関数として感染者数最小化と死亡者数最小化を検討した。また、接種可能なワクチンの全期間にわたる総量が制限されているほか、供給速度にも制限があることを仮定している。目的関数やパラメータの違いにより最適戦略が変化することを確認した。季節性、新型、季節性の順に接種を切り替えるなど、非自明な動的変化が見られた。

#### 研究テーマ E 「新型・季節型同時流行の複雑ネットワークモデルにおけるワクチン接種」

本テーマにおいては、前テーマと同様に新型インフルエンザと季節性インフルエンザの同時流行を想定し、さらに接触の多様性を考慮した複雑ネットワークモデルの上で、ワクチンの最適な配分割合について検討した。他のパラメータが共通であっても、ネットワーク構造が違うだけで最適な配分割合が大きく変化することが確認された。

#### 研究テーマ F 「同時流行時のクラスタによる影響の解析」

ネットワーク上を伝播する感染症の挙動において、クラスタが及ぼす影響は重要な問題である。本テーマにおいては、既存のクラスターネットワークモデルを任意の大きさのクリークに一般化したモデルを提案し、クラスタサイズの分布、巨大クラスタのサイズ、相転移点などの特徴量を導出した(論文 3)。また、このクリークランダムネットワーク上のボンドパーコレーションを考え、浸透確率がクリークによって異なる場合を考えた。これを二つの相互作用する感染症に応用することにより、クラスタ構造が感染症の挙動に影響を与えることを示した。また、安定性の解析により、このモデルにおけるネットワーク構造と感染症伝播との関係を明らかにした。

#### 研究テーマ G 「コストを考慮した複雑ネットワーク上のワクチン接種最適戦略」

ワクチン接種にはコストがかかるため、医療コストとはトレードオフ関係にある。本テーマにおいては、スケールフリーネットワーク上の SIR モデルにおいてコストを考慮したときの最適な接種戦略を検討した。平均場近似による微分方程式モデルを用いて最適化を行った結果、供給可能なワクチンの量や治療コストなどの設定により、必ずしも次数の高いノードから配分するのではなく、非自明な配分が最適となり得ることを示した。

#### 研究テーマ H 「危険度情報に基づく人の移動を考慮したメタポピュレーションモデル」

本テーマにおいては、感染症の感染拡大に関する情報が人々の行動に与える影響の解析を行った(論文 4)。多数の国や都市の間を人々が行き来する状況をネットワーク構造を用いて抽象化して表現した感染症モデルであるメタポピュレーションモデルをベースに、人々の行動の変化を考慮した感染症の数理モデルを構築した。このモデルでは、人々は各地域の感染状況の情報を得て、それに基づき行動を変化させる。このモデルに対して、複雑ネットワーク理論の平均場近似手法を用いることにより、基本再生産数等を理論的に導出した。さらに、この結果を複雑ネットワークモデルおよび日本の国内航空ネットワークの実データに適用して、人々の行動の変化が感染伝播に及ぼす影響を解析した。具体的には、人々が感染の拡大している地域と拡大していない地域のどちらを選好するか表現するパラメータを変化させたときに、感染が拡大するために必要な感染力と、感染が拡大する範囲を評価した。その結果、人々が感染の拡大していない地域を選好すると、低い感染力でも感染拡大を引き起こし、広い範囲に感染が広がることを明らかにした。この結果は、現実の感染拡大においても、人々が自身の安全を考えて、感染が拡大している地域を避け、感染が拡大していない地域を選んで訪れることが、全体としては感染の拡大を促進し得ることを示唆している。

### 3. 今後の展開

感染症の多様性を考慮した数理モデル研究と、データ同化研究の融合をさらに進めていきたいと考えている。同時に感染症に対する対策に活用するためには、より実用的なシステム構築に向けた研究が必要であると考えられる。また、パーソントリップ調査などのデータをベースに、人の移動のモデリングに関しては継続的に新しい数理的アプローチを探っていきたい。さらに、情報伝播など感染症と類似の数理的構造を持つ現象においても、本研究の成果を基

盤に、データ同化等による数理的アプローチの可能性を考えていきたい。

#### 4. 自己評価

当初はパーソントリップ調査データを有効活用できるような人の移動の数理モデルの構築し、それを組み込んだ感染症モデルを構築することを目指して研究を進めていたが、最終的にはメタポピュレーションモデルなどによる既存の感染症モデルに基づく研究となってしまったことが心残りである。また、感染症の多様性を考慮した数理モデル研究と、データ同化研究とをまとめきれなかったことも心残りである。この2点を除いては、当初想定していた形とは違った形ではあるものの感染症伝播のデータ同化シミュレーションに関する成果が得られたほか、理論モデルの研究に関しては当初の想定以上の成果が得られたと考えている。本研究テーマは、研究成果の社会還元を考えることも重要であることから、研究期間の後半で得られた成果の発表をさらに進めていきたいと考えている。

#### 5. 研究総括の見解

データ同化技術の導入でインフルエンザ感染伝播のモデルを構築し、解析・予測を行うための数理的基盤技術を開発するという研究であった。

結果的にはメタポピュレーションモデルなどによる既存の感染症モデルに基づく研究となった。また、感染症の多様性を考慮した数理モデル研究と、データ同化研究とをまとめきれしていないようであるが、当初の予定であった感染症伝播のデータ同化シミュレーションに関する成果を得ており、また、理論モデルの研究に関しては当初の予定以上の成果を得ている。

今後、感染症の多様性を考慮した数理モデル研究と、データ同化研究の融合をさらに進めてほしい。

#### 6. 主な研究成果リスト

##### (1) 論文(原著論文)発表

- |   |
|---|
| 1. 江島啓介, 鈴木秀幸, 合原一幸, 東京都市圏パーソントリップ調査データに基づく新型インフルエンザ感染伝播の数理モデリング, 運輸と経済 70, 54-62, 2010.  |
| 2. B. Wang, L. Cao, H. Suzuki, K. Aihara, "Epidemic spread in adaptive networks with multitype agents", Journal of Physics A 44, 035101, 2011.                        |
| 3. B. Wang, L. Cao, H. Suzuki, K. Aihara, "Impacts of clustering on interacting epidemics", Journal of Theoretical Biology 304, 121-130, 2012.                        |
| 4. B. Wang, L. Cao, H. Suzuki, K. Aihara, "Safety-information-driven human mobility patterns with metapopulation epidemic dynamics", Scientific Reports 2, 887, 2012. |

##### (2) 特許出願

研究期間累積件数: 0 件

##### (3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

・江島啓介, 鈴木秀幸, 合原一幸, "パーソントリップ調査データに基づくインフルエンザ感染

- 伝播の数理モデリング”, 土木学会土木計画学研究発表会, 2010年6月.
- 江島啓介, 鈴木秀幸, 合原一幸, “パーソントリップ調査データに基づくインフルエンザワクチン接種戦略の評価”, 日本数理生物学会年会, 2010年9月.
  - B. Wang, L. Cao, H. Suzuki, K. Aihara, “Epidemic spread in adaptive networks with multitype agents”, Dynamics Days Asia Pacific, July 2010.
  - B. Wang, L. Cao, H. Suzuki, K. Aihara, “Bond percolation on clique random networks and their applications to arbitrary interacting epidemics”, Int. Symp. Innovative Mathematical Modelling, February 2011.
  - B. Wang, L. Cao, H. Suzuki, K. Aihara, “Bond percolation on clique random networks and their applications to arbitrary interacting epidemics”, Int. Conf. Complex Systems, June 2011.
  - B. Wang, L. Cao, H. Suzuki, K. Aihara, “Safeness-information-based human mobility patterns for disease metapopulations”, Int Symp. Innovative Mathematical Modelling, May 2012.

# 研究報告書

## 「金融市場における相転移の時空間構造の自動抽出と予測」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 高田 輝子

### 1. 研究のねらい

金融バブルは世界中で繰り返し発生し、その崩壊時には経済・社会に深刻なダメージをもたらしてきている。従って、その制御のためのメカニズム解明や暴落時期の予測は社会的緊急課題である。金融バブルは相転移現象としてとらえることができる。相転移現象は、発生頻度が低いにも関わらず異常値等により扱いにくく、抽出できる情報量が少ない。しかしその一方で、相転移現象を生み出すシステムは多くの因子を含む非常に複雑なものである。本研究では、大規模データを利用して情報量を増大することに加え、データからのロバスト・効率的データ抽出のための統計手法を開発することにより、少ない情報で複雑なシステムを説明する困難さの克服を目指す。そして相転移現象の基本的事実やパターンを明らかにし、予測や制御に役立てることが、本研究のねらいである。

### 2. 研究成果

#### (1) 概要

本研究は、二つの方向から相転移解析の難しさの克服を図った。金融市場における相転移解析のためのロバスト・効率的情報抽出を可能にする新しい統計手法の開発(研究テーマ A)の方向と、大規模データ利用による情報量増大の方向である。大規模データ利用は、相転移予測精度向上にも寄与したが、中心的成果は金融バブル生成・崩壊メカニズム解明において核心的な因子である投資家心理についての大規模データを用いた新事実発見(研究テーマ B)であった。

研究テーマ A としては以下の3つの点について新しい手法の開発を行い、データからの情報抽出の効率化を実現した: ①相転移予測前兆指標のロバスト化と精度向上を実現する新手法の開発、②SVM を用いた相分類による相転移予測精度向上、③金融バブルのスケール標準化法の提案による主要相転移前兆指標についての情報量増大の実現である。これら成果により、研究開始当初想定していた本研究進展にあたっての技術的困難は概ね解消された。その結果、従来法を用いた場合よりも高精度な相転移予測を実現した。

研究テーマ B としては金融バブル期の投資家心理に関わる解析を中心に行い、以下の結果を得た: ①金融バブル生成・崩壊期における売手/買手別投資家心理の非対称性の可視化と、②ウェブ上の検索行動の株価暴落予兆性の発見、③投資家ネットワークの形状パターンと金融バブルの関係分析や、金融政策の投資家心理への影響分析である。投資家の行動や心理に関わる事実解明は、金融バブル拡大の事前抑制や暴落後の市場鎮静化のためのシステム・デザインに役立つ重要なものである。

#### (2) 詳細

本研究では、いくつかの新しい統計手法の開発により、ロバスト・効率的データ抽出を実現した上で、特に投資家心理に注目した大規模データ解析を行い、いくつかの新事実発見を行った。

**【研究テーマ A: 相転移解析におけるロバスト・効率的情報抽出の為の新しい統計手法開発】**

以下の3点について新しい手法の開発を行い、データからの情報抽出の効率化を実現した:

**① 相転移予測前兆指標のロバスト化と精度向上を実現する新手法の開発**

生態系や気象の分野などを中心に、現在利用されている相転移予測の前兆指標としては、分散、系列相関度(AR1)、歪度、要素間の空間的パターン出現などが、主要なものとして挙げられる。これらをロバスト化・効率化して金融市場に応用するために、本研究ではまず、分散や歪度について、裾厚分布に強いノンパラメトリック確率密度推計値ベースの統計値を提案した。系列相関度については、pointwise mutual information ベースの瞬間的相関度を開発し、ロバスト化・効率化を実現した(論文1)。これは AR1 と異なり、非線形相関度を局所時点で計測できる画期的な方法である。要素間の空間的パターンの例としては、微生物の増殖の相転移の前に出現する特徴的なパターンなどがある。これを金融バブルに応用するにあたり、本研究では投資家同士のネットワーク・パターンの時系列推移を、1998 年以降直近までの米国ダウ30株価指数構成銘柄についての投資家用掲示板データを用いてまず可視化した。そしてそれをマイクロベースの投資家間相互作用度を表す指標に集計し、「市場が一方向に向かうときに投資家間相互作用度が高まる」傾向と、その相転移前兆指標としての有効性を示した(論文投稿準備中)。投資家間相互作用度は、マクロデータをベースにして計測されるものが通例であり、真にマイクロベースで広範囲に渡って投資家間相互作用度を計測したのは、本研究が初めてである。これら前兆指標の提案法は生態系や気象など、他の複雑系分野にも影響しうる貢献である。

**② 相分類に SVM を導入することによる相転移予測精度向上**

金融市場における上昇期と下落期の特徴を、価格収益率データのモーメント情報から SVM に逐次学習させ、上昇期、移行期、下落期の統計的特徴を分析した。また学習結果を基に将来の市場が上昇傾向か下落傾向かを SVM に予測させ、市場平均を大きく上回る投資パフォーマンスを実現させた(その他の成果1)。モーメント情報の代わりに有意な相転移前兆指標を利用すれば、更なる予測精度向上も期待できる。

**③ 金融バブルのスケール標準化法の提案による情報量増大の実現**

金融バブルは数十年に 1 度の頻度でしか発生しないため、現在利用可能な統計データには限りがあり、十分な解析を行いにくい。しかし、地震の場合と同様に、スケールの小さいミニバブルであれば発生頻度も高く、十分な統計量が確保できる。本研究では、金融バブルのスケールの標準化をボラティリティーについて行う新しい手法を提案し、相転移予測に有益ないくつかの新事実を発見した(その他の成果3)。ボラティリティーは、相転移予測の前兆指標の中でもシステム安定性との関係が特に深い重要指標であり、これについての情報量増大手法の提案と新事実発見は、他分野の相転移解析にとっても重要な貢献であると考えられる。



以上の新しい統計手法開発の成果により、研究開始当初想定していた本研究進展にあたっての統計手法面の技術的困難を、概ね解消することができた。その結果、従来法を用いた場合よりも高精度な「金融市場における相転移予測」を実現させることができた(論文投稿準備中)。

#### 【研究テーマB:大規模データによる投資家心理についての新事実発見】

金融バブル期の投資家心理に関わる統計解析を行い、以下の結果を得た:

##### ① 金融バブル生成・崩壊期における売手/買手別投資家心理の非対称性の可視化

ニューヨーク証券取引所全上場銘柄について、利用可能最長株式の指値注文表の気配値データを用いて、金融バブル生成・崩壊局面における売手/買手別不安度の時系列推移パターンの非対称性を可視化した。その結果、「金融バブル生成局面においては非対称度が拡大し、金融バブル崩壊後はそれが解消し対称度が高まる」という事実をデータで示すことができた(論文投稿準備中)。これだけ大規模な板値データ解析はほとんど例が無い。また「対称性の破れ」は物理学において、一般的な相転移に関係の深い現象であり、この発見は、金融バブルも普遍的な物理法則に支配されている可能性につながりうるものといえる。

##### ② ウェブ上の検索行動の株価暴落予兆性の発見

金融バブルの生成過程で重要な役割を果たすのは機関投資家よりもむしろ一般投資家である。そしてその一般投資家は、実際に投資行動を起こす前に、インターネット上で情報を確認してから投資判断を下すことが多いと考えられる。本研究では、ウェブ上の検索行動と株価変動の関係を各国主要株価指数について分析した結果、株価の上昇時/下落時における検索行動の非対称性や、検索頻度統計による株価暴落予測可能性を発見した(その他の成果2)。投資行動の方向性決定の前段階である検索行動と、投資行動の結果である株価変動との関連性を分析した研究はまだ非常に少ない。

##### ③ 解析途上の研究

###### 投資家ネットワーク分析

収集した投資家ネットワーク時系列データの、株価変動とネットワーク形状パターン変動の関係についての分析を行った。これだけ長期の大規模な社会ネットワークの解析は例が無く、学術的価値が高いものである。

###### 金融政策が投資家心理に与える影響分析

1990年代初め頃からの格付けが利用できる米国上場のほぼ全ての社債の長期時系列データを用いて、金融政策が投資家のリスク選好度(投資意欲)にどのように影響を与えているかについての分析を行い、いくつかの新事実を発見した。本研究は金融政策から投資家心理へのチャンネルをつなげるものであり、金融政策デザインのために不可欠なものである。

金融バブル生成・崩壊期における投資家の行動や心理に関わる事実解明は、投資家間相互

作用を通じて価格上昇/下落期待が高まる過程という核心部分の解明につながるものであり、最終的には金融バブル拡大の事前抑制や暴落後の市場鎮静化のためのシステム・デザインに役立つ、重要なものである。

### 3. 今後の展開

今後の展開については、大きく分けて二つの方向を考えている。一つは、生態系や物理系など他の複雑系における相転移現象と金融バブルとの間の共通パターンに注目し、一般的な相転移現象について既に得られている予測やメカニズムについての知見を活用し、金融バブル研究に応用していく方向である。もう一つは、金融バブルの生成・崩壊の核心的因子である投資家心理のダイナミクスを群集行動としてとらえ、金融市場の高頻度数値データだけでなく、大規模ウェブ・テキストデータなども用いて、ミクロな投資家レベルからマクロな株価変動レベルへの伝播過程を観察し、基本的事実発見、メカニズム解明、制御を目指す方向である。今後はこの二つの方向から、金融バブルの真実に迫り、社会に役立てる研究成果に結実させていきたい。

### 4. 自己評価

本研究の当初目標は、金融市場における相分類による相転移予測を、大規模金融データとロバスト・効率的手法の開発により、実用レベルの精度に高めることであった。さきがけ研究の成果は、相転移研究のための技術的困難を、統計手法とデータの大規模化の両面から、概ね解消することができたことである。統計手法面では投資家間相互作用度の開発も、ミニバブル情報の発掘による金融バブル情報増大も、さきがけ期間に着手できるとさえ想定していなかった。また、データの大規模化による情報量増大面についても、当初予定の日米株式逐次データのみならず、日米の売手/買手別指値表による板値の最長時系列データや、ウェブ上の大規模投資家掲示板最長時系列データ、格付けが利用できる米国上場のほぼ全ての社債データなど、異なる種類について全テキストファイルで約280TBに達するなど、比類の無いレベルの大規模データを収集・クリーニングし、解析を行うことができた。

一方、本研究の技術的困難をさきがけ期間中に解決することを目指したことに加え、データ・クリーニングや大規模データ用クラスター・マシンの整備といった「大規模データ解析環境整備」に想定以上に時間がかかってしまった結果、成果が総花的となってしまう、着手した研究を全てさきがけ期間内にまとめきることができなかった点は、目標通りだったとは言えない。できるだけ早いタイミングで全て完成させたい。さきがけのおかげで、当初想定 of 技術的困難はクリアされ、計算環境も大規模データも整備された。しかも、今後本研究を発展させていく上で発生しうる技術的困難点については、共同研究などにより、多くのさきがけ研究者やアドバイザーの先生方からの協力を得られる環境まで整ったことは、心強い限りである。あとは、ロバスト・効率的な大規模データ解析により、この研究を大きく花開かせるよう、努力していきたい。

### 5. 研究総括の見解

金融・経済データ情報から金融バブルなどの相転移現象の時空間構造を可視化し、相転移メカニズムの探索・モデル化と予測を可能にする知識生産方法を開発し、金融・経済のシミュレーションモデルの新しいデータ同化技術構築の基盤技術を目指すという研究課題である。

当初目標は、金融市場における相分類による相転移予測を大規模金融データとロバスト・効率



的手法の開発により、実用レベルの精度に高めることであったが、相転移研究のための技術的困難を、統計手法とデータの大規模化の両面から、ほぼ解消することができている。また、さきがけの特徴である研究者間のコラボレーションをうまく活用し、当初の研究範囲を超え、日米の売手/買手別指値表による板値の最長時系列データやウェブ上の大規模投資家掲示板最長時系列データ、格付けが利用できる米国上場のほぼ全ての社債データ等々、他種の大規模データの収集・解析を行い、相転移予測の精度を高めることができたことを評価する。

今後、他の複雑系における相転移現象で得られている知見の活用や、投資家心理のダイナミクスの分析等を進め、金融バブルの真実に迫り、社会に役立つ研究成果に結実させて行ってほしい。

## 6. 主な研究成果リスト

### (1) 論文(原著論文)発表

1. Teruko Takada. Mining Local and Tail Dependence Structures based on Pointwise Mutual Information. *Data Mining and Knowledge Discovery*. 2012, 24(1), pp.78-102.

### (2) 特許出願

研究期間累積件数:0件

### (3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. Teruko Takada and Takahiro Kitajima. Phase Classification by Support Vector Machine. *Computational Statistics and Data Analysis International Conference on Computational and Financial Econometrics*, 2010.
2. 高田輝子、佐藤圭. ウェブ検索頻度による投資家不安度の分析. 日本統計学会第80回大会、2012.
3. Teruko Takada and Akimitsu Inoue. Multiple Time Scale Volatility Patterns before Abrupt Switching in Financial Markets. *Computational Statistics and Data Analysis International Conference on Computational and Financial Econometrics*, 2012.
4. 高田輝子. 金融バブルのような相転移現象の予測法開発. 競創ダイナミクスの統合的理 解 第1回研究会, 2013. (招待講演)
5. 高田輝子. 金融バブルの大規模データ解析. 第27回人工知能学会全国大会オーガナイズド・セッション「金融情報学」, 2013. (招待講演)

# 研究報告書

## 「空間的な情報システムの設計開発支援システム」

研究タイプ: 通常型

研究期間: 平成 21 年 9 月～平成 25 年 3 月

研究者: 中西泰人

### 1. 研究のねらい

ディスプレイやプロジェクタなどの情報出力装置の性能の向上と価格の低下と共に、複数の入出力装置が空間的に配置されるようになり、電子広告やプロジェクションマッピング等の新たな映像空間体験が都市空間および建築空間に広まりつつある。こうした情報システムを構築する際には、実空間の中で入出力装置をどのように空間的に配置するか、そしてそこに表示する情報の内容や構成をどのように設計するかを、同時に考える必要がある。しかしながら、そうした空間デザインと情報デザインを総合的に取り扱えるような設計開発支援システムは存在していない。そのため、配置しやすい場所に入出力装置が置かれ、空間の特徴を考慮することなく他のメディアの為にデザインされた情報がそのまま提示されることが多いという問題点がある。

そこで本研究では、提示する情報の大きさや動き、出力装置のサイズや入力装置と出力装置の配置関係、周囲の空間を本質的な要素とする情報システムを空間的な情報システムと呼び、その設計開発を支援すべく、下記の機能を備えた設計開発支援システムを構築することを研究の目的とする: 1) 簡易な空間モデリングソフトとソフトウェア統合開発環境を併用できる、2) モデリングした仮想空間の中でインタラクティブな情報システムが実行される様子をシミュレーションできる、3) シミュレーションの実行中に出力装置の位置と向きおよび入力装置の位置と向きを変更できる。またこの設計開発支援システムの実際的な利用を通じて、設計開発を行うプロセスを詳細化し、求められる新たな機能への要求を探り実装するサイクルを繰り返す。そしてこのシステムを利用するユーザ達の知が結節する媒体となり得るよう、製作過程の表記法を開発することで設計開発の支援を行う。これらの研究項目を通じて情報空間と実空間とを総合した環境をデザインする基盤的な技術となることを目指す。

### 2. 研究成果

#### (1) 概要

上で述べた本研究の狙いをサブテーマに分割し、簡潔な言葉で言い換えると、(1) つくる道具をつくる、(2) つくる過程を提案し分析する、(3) つくる過程の表記方法をつくる、となる(図 1)。要素の組み合わせり方がこれまでにないシステムを作るための道具は、新たな道具である。道具が変われば製作過程じたいにも変化がおき、アイデアを外在化する過程、ふくらませる過程、検証する過程も変わる。そのため本研究では、(1)と(2)をデザイン思想的に反復的に進めながら、その過程および成果物を分析した結果を(3)へとフィードバックするという研究の進め方をとった。つくる道具としては、オープンソースの 3D ゲームエンジンである jMonkeyEngine(<http://jmonkeyengine.org>)を拡張し、ビジュアルプログラミングのプロトタイプツールとして広く活用されている Processing(<http://processing.org>)の実行画面が 3D 仮想空間

内に表示できるライブラリである CityCompiler を構築した。つくる過程としては、仮想空間を用いたシミュレーションに加え、仮想空間と模型空間を併用したシミュレーションを提案し、それらの比較を行った。つくる過程の表記伝達方法としては、中島・藤井・諏訪らによるイノベーションプロセスのモデルである FNS モデルを拡張したものを用いた。以下に、個々のサブテーマについてより詳細に説明する。

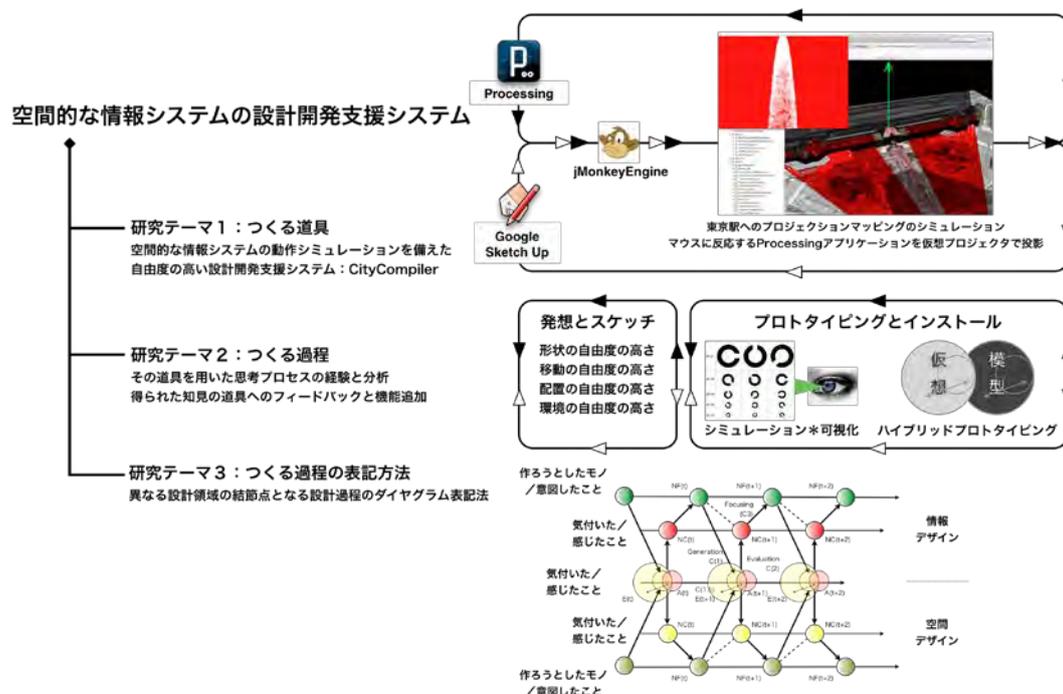


図1 本研究のサブテーマおよび成果の概要

## (2) 詳細

### 研究テーマ1)

本研究では、仮想空間でシミュレーションするだけでなく、文字や映像などグラフィックスを表示するコードはそのまま実空間で利活用できるよう、Processing の実行画面がそのまま表示できるシミュレータとして設計開発支援システムを構築した。仮想ディスプレイおよび仮想プロジェクタを実装するにあたっては、Processing がグラフィックスを描画する度にテキストを生成して 3D ゲームエンジンに渡すよう実装を行った。また仮想空間内の様子をセンシングする仮想カメラや仮想センサの実装も行うにあたっては、実カメラや実センサと同じ取り扱いができる実装とした。これにより、仮想空間における動作であるか実空間における動作であるかを表すひとつの変数の値を切り替えるだけで、シミュレーションの後に実空間での運用と配置を行うことができる枠組みを提供した。

構築した設計開発支援システムのソースコードおよびドキュメントは下記の URL <https://github.com/yasutonakanishi/CC4p52> で公開している。

### 研究テーマ2)

カメラやセンサを入力とする Processing アプリケーションは空間的な情報システムの構築に広く用いられているが、パーソナルコンピュータのスケールでは開発が困難な大型の出力装置や複数の入出力装置を用いた情報システムを設計開発することには困難が伴う。

そうした物理的な配置の困難さが発想や実装の制約となる場合でも、仮想空間におけるシミュレーションを用いれば自由にアイデアを試すことができる。制約がむしろ発想の飛躍をもたらす場合もあるが、ユーザらによる CityCompiler を使った製作過程からは、1) 形状の自由度の高さ(大きさ, 縦横比, 複雑な形状)、2) 移動の自由度の高さ(高速に移動可能, 配線がない)、3) 配置の自由度の高さ(水上や水中や高所, 複数の装置の複雑な配置)、4) 環境の自由度の高さ(有名建築等の特異な形状, 周辺の空間の文脈)、といった自由度の高さが発散的思考の広がりをもたらしていたことが見て取れた。

そしてさらに、入出力装置を仮想空間のものと実空間のものとをスムーズに切り替えることができる枠組みがもたらした新たな製作過程として、仮想空間と模型空間を組み合わせるハイブリッドプロトタイピングを提案した。同じ処理を行う情報システムを仮想空間と模型空間で動作させることで、仮想空間では発見できない問題点を模型空間で発見できるという利点がある。仮想空間と模型空間を組み合わせる方法として、情報システムを動作させる空間を交互に切り替える方法と、複数の入出力装置のうち一部を仮想空間内のものを用い一部は模型空間のものを混合する方法の 2 つを提案し、それらの長所と短所を考察した。また仮想空間と模型空間を交互に切替える方法と仮想空間に可視化を組み合わせた方法との比較を行い、それらの長所と短所を考察した。

これらの結果を総合することで、発散的思考を通じてアイデアを出すフェイズでは仮想空間を用い、現実的な実装を確認しながらプロトタイピングするフェイズでは仮想空間に可視化を織り込みながら模型空間を混合させ、実空間への配置を前提にシステムをチューニングし収束させるプロトタイピングには仮想空間と模型空間を切替えて用いる、という過程を空間的な情報システムを設計開発するひとつの方略として提案した。

### 研究テーマ 3)

空間的な情報システムにおいては、同じソースコードを用いたシステムであっても、機器の形状やサイズ、機器同士の位置関係や空間の中の配置等によっては、現象する映像や意味が異なるものになることが多い。そのため本研究ではシミュレータを構築しているが、さらに製作した過程を表記する方法があれば、自らの思考過程を外在化しやすくなると共に他者とそれを共有できる。本研究ではその表記方法として FNS モデルを情報デザインと空間デザインを並行的に進めるモデルに拡張した図を用いた。Web を通じて参加者を募ったワークショップにおいて、参加者が自らのシステムを発表する際にその製作過程を記述し伝達する際にこの図を用い、その有効性を検討した。



ワークショップ第 1～3 回の様子および参加者らによる製作過程の記録

### 3. 今後の展開

また本研究で設計開発の対象とできた出力装置は映像装置のみであった。音や他のユーザの身体動作といった他のコミュニケーションモードを設計開発の対象とできるよう、システムを拡張していきたい。空間的な情報システムの大きな特徴として、複数人のユーザを内包する環境としてそれが機能するということがある。そのため設計すべきインタラクションは、人と環境とのインタラクションだけではなく、人と人とのインタラクションもその対象となる。しかしながら本研究では、人と人とのインタラクションを設計開発の対象にはできなかった。そのためシステムを実空間に配置して現象する人と環境および人と人とのインタラクションを計測し、その結果を踏まえてシステムを再構築する過程も支援の対象としたい。

### 4. 自己評価

このさきがけ研究期間の中で、空間的な情報システムを設計開発する方法を複数提案してそれらの長所短所を考察しながら、設計開発支援システムに機能を追加しバージョンアップを繰り返すことができた。本システムを用いて開発した空間的な情報システムの実運用は美術館の中と研究場所近辺の屋外だけに留まっており、広汎な都市空間を対象とした空間的な情報システムの運用には至らなかったことは残念である。しかしながら最終年度にはユーザを一般公募してワークショップを開催し、広く利用してもらうためのドキュメント整備も行うことができた。研究を通じて新たな知見を得ることに加えて、自由度の高い設計開発支援システムをすぐに利用可能なかたちとして開発したことは、評価できると考えている。

### 5. 研究総括の見解

デザインプロセスをモデル化し、それに基づいた設計支援システムの構築が課題である。

研究期間中、空間的な情報システムを設計開発する方法を複数提案し、適用・評価し、設計開発支援システムに機能を追加してバージョンアップを繰り返している。広汎な都市空間を対象とした空間的な情報システムの運用には至らなかったとはいえ、ユーザを一般公募してワークショップを開催し、ドキュメント整備も行い、こんごの普及体制を整えたことを評価したい。

今後、映像装置だけでなく、他のコミュニケーションモードを設計開発の対象とできるようシステムを拡張し、広汎な都市空間を対象とした空間的な情報システムに取り組んでほしい。

### 6. 主な研究成果リスト

#### (1) 論文(原著論文)発表

1. 中西泰人, アーキテクチャとインタラクションデザイン, 情報処理学会誌, Vol. 51, No.7, pp.759-766, 2010.
2. Yasuto Nakanishi, Koji Sekiguchi, Soh Kitahara, Takuro Ohmori, Daisuke Akatsuka, Hybrid prototyping by using virtual and miniature simulation for designing spatial interactive information systems, Proceedings of the 9th international conference on Pervasive computing, pp. 250-257, Springer-Verlag, 2011.
3. Yasuto Nakanishi, Koji Sekiguchi, Takuro Ohmori, Soh Kitahara, Daisuke Akatsuka, Roles of miniature space in hybrid prototyping, Proceedings of the 8th ACM conference on

Creativity and cognition, pp.379-380, ACM, 2011.

4. Yasuto Nakanishi, Virtual prototyping using miniature model and visualization for interactive public displays, Proceedings of the Designing Interactive Systems Conference, pp.458-467, ACM, 2012

(2)特許出願

なし

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. 平野啓一郎、森野和馬、ケンイシイと共同制作「DAWN」、東京都現代美術館に展示(2010年2月～3月)
2. 平野啓一郎と共同制作「記憶の告白 - reflexive reading」、佐世保市島瀬美術センターに展示(2011年7月～9月)
3. 想像性と創造活動を広げるシステムを作りたい JST news 2012-9月号, P.16
4. 情報・生活を豊かで楽しくするための作り方, NHK ラジオ第一, 私も一言! 夕方ニュース「ここに注目」, 2012年11月6日出演
5. ワークショップ「空間をプログラミングしよう!」第1回開催(2012年11月24日)
6. ワークショップ「空間をプログラミングしよう!」第2回開催(2012年11月25日)
7. ワークショップ「空間をプログラミングしよう!」第3回開催(2013年1月19日)
8. 平野啓一郎と共同制作「記憶の告白 - reflexive reading」、高知県立文学館に展示(2013年2月～4月)